

09/965594

**STIC-Biotech/ChemLib**

---

**From:** Schnizer, Richard  
**Sent:** Wednesday, August 27, 2003 2:28 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/965,594

Please search the commercial and published application databases for polypeptide SEQ ID NOS: 1, 12, 14, 16, 18, 20, 22, and 26 from 09/965,594. Please also search for nucleic acids that could encode these polypeptides.

Thank you-

Richard Schnizer, Ph.D.  
Patent Examiner  
Art Unit 1635  
CM1 12E17  
703-306-5441  
Mail Box CM1 11E12

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:18:33 : Search time 2365.6 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 NAPIYAAQOTRGLGCIIT.....GVAKAVDFIPVLSLETIMRS 182

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2.1/uspto.spool/US09965594/runat\_29082003\_151919\_28310/app.query.fasta.i.2872  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MFN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594@cgn.1.14686@runat\_29082003\_151919\_28310 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*

1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.ua:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pln:\*  
35: em.htg\_rod:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	946	99.3	2058	6	AX395309 Sequence
2	946	99.3	2058	6	AX454818 Sequence
3	945	99.2	543	14	AF369218 Hepatitis
4	945	99.2	543	14	AF369235 Hepatitis
5	943	99.0	5360	6	AR118686 Sequence
6	943	99.0	5360	6	I06434 Sequence 48
7	943	99.0	5360	6	I09328 Sequence 8
8	943	99.0	6785	6	AR118692 Sequence
9	943	99.0	6785	6	I06440 Sequence 54
10	943	99.0	6785	6	I09329 Sequence 10
11	943	99.0	7310	6	AR118696 Sequence
12	943	99.0	7310	6	I09331 Sequence 15
13	943	99.0	7310	14	HPCPOLYP
14	943	99.0	8316	6	AR118703 Sequence
15	943	99.0	8987	6	AR118728 Sequence
16	943	99.0	9185	6	AR118722 Sequence
17	943	99.0	9185	6	AR118723 Sequence
18	943	99.0	9185	6	BD091382 HCV culti
19	943	99.0	9185	6	I08294 Sequence 1
20	943	99.0	9379	6	AR166930 Sequence
21	943	99.0	9379	6	AR301300 Sequence
22	943	99.0	9401	6	AR176483 Sequence
23	943	99.0	9401	6	BD080334 Hepatitis
24	943	99.0	9401	6	E66593 Hepatitis C
25	943	99.0	9401	6	I71894 Sequence 9
26	943	99.0	9401	6	I81885 Sequence 9
27	943	99.0	9401	14	HPCPLYPRE
28	943	99.0	9609	12	AF387805 Synthetic
29	943	99.0	9609	12	AF387808 Synthetic
30	943	99.0	9618	14	AF271632 Hepatitis
31	943	99.0	9646	12	AF387806 Synthetic
32	943	99.0	9693	12	AF387807 Synthetic
33	942	98.8	543	14	AF369222 Hepatitis
34	942	98.8	543	14	AF369232 Hepatitis
35	942	98.8	543	14	AF369240 Hepatitis
36	942	98.8	543	14	AF369245 Hepatitis
37	942	98.8	2061	6	AX441176 Sequence
38	942	98.8	2061	6	AX467113 Sequence
39	940	98.6	543	14	AF369224 Hepatitis
40	940	98.6	543	14	AF369237 Hepatitis
41	940	98.6	9424	14	AF511948 Hepatitis
42	939	98.5	543	14	AF369230 Hepatitis
43	939	98.5	9502	6	E08263 grNA of Hep
44	939	98.5	9502	6	E08264 cdNA of Hep
45	939	98.5	9502	14	HPCRCUJ

# ALIGNMENTS

RESULT 1



```

AX395309
LOCUS AX395309 2058 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 2 from Patent WO0196875.
ACCESSION AX395309
VERSION AX395309.1 GI:21066308
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,L.C.,
Colt,D. and Medina-Selby,A.
TITLE Hcv antigen/antibody combination assay
JOURNAL Patent: WO 0196875-A 2 20-DEC-2001;
CHIRON CORPORATION (US)
FEATURES
source
1..2058
Location/Qualifiers
1..2058
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="representative NS3/4a conformational antigen"
CDS
1..>2058
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/db_xref="GI:21066309"
/transl_table="MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAOTF
LATCINGVCWTVYHAGCTRTIASPKGPVIONYNVDODLVGWPAQGSRLPTCTGGS
SDLYLVTRHADVPVRRRGSRLSPRPISYLGSSGGLPCPAGHAGVIFRAAIVC
TRGVAKAVDFIPVNLVSVAAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYKFLA
YAAGGYKVLVNLPSVAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYKFLA
DGGCGGAYDIIICDECHSDATSLIGIGTVLQDAETAGARLVLAATAPPGSVTVPH
PNIIEVALSTTGEIPFYGKAIPLEVYKGGRLIFCHSKKCDLAKLVALGINAVAY
YRGDLSVIPPIGDWWVVDALMTGTYGDFSDVIDCNTCVOTDFSLDPTFTIETI
TLPDQAVSRTQRRGTGRGKPGIYRFVAPGERPSGMPDSSVCECCDAGCANWELTPA
ETTVLRAYMNTPTGLPCVDHLEFEGVFTGLTHIDAHFLSQTKQSGENLPYLVAOA
TVCAQAQPPSWDQMKCLIRLPTLHGPTTLYRLGAYONETITHPVTKYIMTCM
SADLEWVTSTWLVGGVLAALAAAYCLSTGCVIVIGRVVLSGKPAIIPDREVLYREFDE
MEEC"
BASE COUNT 419 a 634 c 580 g 425 t
ORIGIN
Alignment Scores:
Pred. No.: 6,56e-68 Length: 2058
Score: 946.00 Matches: 180
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.90% Mismatches: 0
Query Match: 99.27% Indels: 0
DB: 6 Gaps: 0
US-09-965-594-1 (1-182) x AX395309 (1-2058)
QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20
Db 1 ATGGCGCCCATCACCGCTAGCCAGCAGACAGAGCGGCCCTCCTAGCGTCGATACAC 60
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGCGCGCGCAAAAACCAAGTGGAGGTGAGGTCAGATTGCTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValTyrHisGlyAla 60
Db 121 GCCCAAACTTCTGGCAACGTCATCATGAGGCTGCTGACTGCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAACGAGGACCATCGCGTCACCAAGGCTCCTGTGCATCCAGATGATACCAATGAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGGTGGCGCGCTCCGAAAGGTAGCCGATCATTCACACCCCTGCAC 300

```

```

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 301 GGCTCCTCGGACCTTTACCTGGTCACGAGCAGCGCATGTCTCCGTCGCCCGCGCG 360
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 361 GGTATAGCAGCGGACGCTGCTGCGCCCGGCCCATTTTCCTACTTGAAGGCTCCTCG 420
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValCys 160
Db 421 GGGGTCGCTGTTGTGCGCCCGCGGCGACGCGCTGGGCATATTAGGCGCGCGGTGC 480
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCGCTGGATGGCTAAGCGGTGACTTTATCCCTGTGGAGAACCTAGACACCACTG 540
QY 181 ArgSer 182
Db 541 AGGTCC 546
RESULT 2
AX454818
LOCUS AX454818 2058 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 1 from Patent WO0196870.
ACCESSION AX454818
VERSION AX454818.1 GI:21714047
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Chien,D.Y., Arcangel,P., Tandeske,L., George-Nascimento,L.C.,
Colt,D. and Medina-Selby,A.
TITLE Immunoassays for anti-hcv antibodies
JOURNAL Patent: WO 0196870-A 1 20-DEC-2001;
CHIRON CORPORATION (US)
FEATURES
source
1..2058
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="representative NS3/4a conformational antigen"
CDS
1..>2058
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD38232.1"
/db_xref="GI:21714048"
/transl_table="MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVQIVSTAOTF
LATCINGVCWTVYHAGCTRTIASPKGPVIONYNVDODLVGWPAQGSRLPTCTGGS
SDLYLVTRHADVPVRRRGSRLSPRPISYLGSSGGLPCPAGHAGVIFRAAIVC
TRGVAKAVDFIPVNLVSVAAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYKFLA
YAAGGYKVLVNLPSVAATLFGFAYMSKAHGDIPNIRTVGRTITGSPITVYKFLA
DGGCGGAYDIIICDECHSDATSLIGIGTVLQDAETAGARLVLAATAPPGSVTVPH
PNIIEVALSTTGEIPFYGKAIPLEVYKGGRLIFCHSKKCDLAKLVALGINAVAY
YRGDLSVIPPIGDWWVVDALMTGTYGDFSDVIDCNTCVOTDFSLDPTFTIETI
TLPDQAVSRTQRRGTGRGKPGIYRFVAPGERPSGMPDSSVCECCDAGCANWELTPA
ETTVLRAYMNTPTGLPCVDHLEFEGVFTGLTHIDAHFLSQTKQSGENLPYLVAOA
TVCAQAQPPSWDQMKCLIRLPTLHGPTTLYRLGAYONETITHPVTKYIMTCM
SADLEWVTSTWLVGGVLAALAAAYCLSTGCVIVIGRVVLSGKPAIIPDREVLYREFDE
MEEC"
BASE COUNT 419 a 633 c 581 g 425 t
ORIGIN
Alignment Scores:
Pred. No.: 6,56e-68 Length: 2058
Score: 946.00 Matches: 180
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.90% Mismatches: 0
Query Match: 99.27% Indels: 0
DB: 6 Gaps: 0
US-09-965-594-1 (1-182) x AX454818 (1-2058)

```

```

Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1 ATGGCGCCATCAGCGGTACGCCACAGACAGAGGGCCCTCTAGGGTCATATACACC 60
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValcIuGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTCTCAACTGCT 120
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValcCystTrpThrValTyrHisGlyAla 60
Db 121 GCCCAAAACCTTCTCGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGGGCC 180
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAAGGAGGACATCGGTACCAAGGTCCTGTATCCAGATGTATACCAATGTAGAC 240
Qy 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGCTGGCGGCTCCGCAAGGTAGCCCATCATTGACACCTGCACCTTGC 300
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 301 GCTCTCTCGGACCTTACCTGGTCACGAGGACCGCCGATGTATCCCGTCCCGCGCGG 360
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 361 GGTATAGCAGGGCAGCGCTCTGTGCGCGCGGCCATTTCTTACTTGAAGGCTCTCTGC 420
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 421 GGGGTCGCGCTGTGTGCGCGCGCGGCGACCGCGTGGGCATATTAGGGCGCGCGTGTGC 480
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACCAACCTG 540
Qy 181 ArgSer 182
Db 541 AGGTCC 546

RESULT 3
AF369218
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
Genetic diversity and response to IFN of the NS3 protease gene from clinical strains of the hepatitis C virus
Arch. Virol. 147 (7), 1385-1406 (2002)
22105140
PUBMED
12111414
REFERENCE
2 (bases 1 to 543)
Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
AUTHORS
Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.
TITLE
Direct Submission
Submitted (09-APR-2001) Infectious Disease Research, Henry Ford Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI 48202, USA
Location/Qualifiers
1..543
/organism="Hepatitis C virus"
/mol_type="genomic RNA"
/isolate="pt.4"
/db_xref="taxon:11103"
/note="type: 1A"

FEATURES
Source

```

```

CDS
<1..>543
/note="polyprotein"
/codon_start=1
/product="NS3 protease"
/protein_id="AAK54543.1"
/db_xref="GI:14150561"
/translation="APITAYAQQTGRGLGCLITSLGRDKNQVEGEVOIVSTAAQTFU
ATCINGCVQTVYHGAGTGTIASPKGPVIQMTYNDKDLVGPAPQGSRSILTPCTCGSS
DLVYTRHADVLPVPRBGDSRGLSPRISYLGSSGGPLLCPAGHAVGIFRAVCT
RGVAAVDFIPVENLETHRS"
BASE COUNT 113 a 166 c 158 g 106 t
ORIGIN

Alignment Scores:
Pred. No.: 1..89e-68 Length: 543
Score: 945.00 Matches: 180
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 99.45% Mismatches: 0
Query Match: 99.16% Indels: 0
DB: 14 Gaps: 0

US-09-965-594-1 (1-182) x AF369218 (1-543)
Qy 2 AlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21
Db 1 GCACCTATCACGGCATACGCCACAGACAGAGGGCCCTCTAGGGTGTATATACCACT 60
Qy 22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41
Db 61 CTACTGGCGCGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACTGCTGCC 120
Qy 42 GlnThrPheLeuAlaThrCysIleAsnGlyValcCystTrpThrValTyrHisGlyAlaGly 61
Db 121 CAGACTTCTTGGCAACATGCATCAACGGGTATGCTGGACCGTCTACCAAGGGCGCGA 180
Qy 62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLys 81
Db 181 ACAGGACCATTCATCACCAAGGTCCCGTCATCCAGATGTACCAATGTATACAAA 240
Qy 82 AspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 101
Db 241 GATCTTGTGGGTGGCGCGCTCTCAAGGTTCCTCATTCACACCTGCACCTCGGCG 300
Qy 102 SerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGly 121
Db 301 TCTCGGACCTTTACTTGTGTGTCAGAGGCATGCGCATGTCTACCGCGTGGCGCGAGGT 360
Qy 122 AspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGly 141
Db 361 GATAGGAGGGGACGCTGTCTGCGCGCGCGCCATTTCTTACTTGAAGGCTCTCGGG 420
Qy 142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161
Db 421 GGTCCGCTGTGTGCGCGCGCGGACAGCGGTAGGCATATTACAGGCGCGCGGTGTGCACC 480
Qy 162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181
Db 481 CGTGAGGTGCTTAAAGCGGTGGACTTTATCCCGTGGAGAACCTTAGAGACCAACCAITGAGG 540
Qy 182 Ser 182
Db 541 TCC 543

RESULT 4
AF369235
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF369235
Hepatitis C virus pt.1Y NS3 protease gene, partial cds.
AF369235
AF369235.1 GI:14150594
Hepatitis C virus
Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

Hepacivirus.  
 1 (bases 1 to 543)  
 Holland-Staley,C.A., Kovari,L.C., Golenberg,E.M., Pobursky,K.J. and  
 Mayers,D.L.  
 Genetic diversity and response to IFN of the NS3 protease gene from  
 clinical strains of the hepatitis C virus  
 Arch. Virol. 147 (7), 1385-1406 (2002)  
 22105140  
 MEDLINE  
 12111414  
 2 (bases 1 to 543)  
 Holland-Staley,C.A., Kovari,L.C., Golenberg,E. and Mayers,D.L.  
 Direct Submission  
 Submitted (09-APR-2001) Infectious Disease Research, Henry Ford  
 Health Systems, 2799 W. Grand Blvd. Rm 7045 E & R, Detroit, MI  
 48202, USA  
 FEATURES  
 source  
 1..543  
 /organism="Hepatitis C virus"  
 /mol\_type="genomic RNA"  
 /isolate="pt.1v"  
 /db\_xref="taxon:11103"  
 /note="type: 1A"  
 <1..>543  
 /note="polyprotein"  
 /codon\_start=1  
 /product="NS3 protease"  
 /protein\_id="AAK54560.1"  
 /db\_xref="GI:14150595"  
 /translation="APITAYAAQTRGLLCILITSLIGRDXNOVEGEVOIVSTAAQTFL  
 ATCINGCVMTVTHGACRTIATSPKGPVIOMTINVDKDLVGMWPAQGSRLTPTCCSS  
 DILYTRHADVTPVRRRGRSGSLSPRIYLYKSSGGPLICPAGHAGVIFRAAVCT  
 RGVAKAVDFIPVENLETTMRS"  
 BASE COUNT 112 a 169 c 159 g 103 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,896-68 Length: 543  
 Score: 945.00 Matches: 180  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.45% Mismatches: 0  
 Query Match: 99.16% Indels: 0  
 DB: 14 Gaps: 0  
 US-09-965-594-1 (1-182) x AF369235 (1-543)  
 Qy 2 AlaProleThrAlaTyAlaGlnThrArgGlyLeuLeuGlyCysIleIleThrSer 21  
 Db 1 GCACCATCACACATACGCCAGCAGACAAAGGGGCTCTAGGGTGCCATAATCACCAG 60  
 Qy 22 LeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAla 41  
 Db 61 CTACCGCGCGGACAAACCAACCAAGTGGAGGGTGAGGTCCAGATTGTCAACTGCTGCC 120  
 Qy 42 GlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGly 61  
 Db 121 CAACITTTCTTGCACAGTCATCAATGGGGTATGGTCCAGCCCTCTACCAAGCGCCGGA 180  
 Qy 62 ThrArgThrIleAlaSerProLysGlyProValIleGlnMetThrThrAsnValAspLys 81  
 Db 181 ACAGGACCATCGATCACCACAAAGGGTCCCGTTATCCAGATTGTACACCAATGTAGACAAA 240  
 Qy 82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrThrProCysThrCysGly 101  
 Db 241 GATCTTGTGGGTGGCGCGCTCTCAAGGTTCCCGCTCAITGACACCCCTGCACCTCGCGG 300  
 Qy 102 SerSerAspLeuThrLeuValThrArgHisAlaAspValIleCProValArgArgGly 121  
 Db 301 TCCTCGACCTTTACCTGGTGCAGAGCAGCGCGATGATCCCGTGCSCCGCGGGT 360  
 Qy 122 AspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerGly 141  
 Db 361 GATAGCAGGGGACGCTGTCTCGCGCGGCCCATTTCTACTTGAAGGCTCCTCGGGG 420

Qy 142 GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThr 161  
 Db 421 GGTCCTCTTTGGCCCGCGGACACCGGTAGGCATATTACAGGCGCGGTGTGCACC 480  
 Qy 162 ArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetArg 181  
 Db 481 CGTGAGTGGCTAAGCGGTGGACITTTATCCCTCTGGAGAACCTAGACACCAACCATGAGG 540  
 Qy 182 Ser 182  
 Db 541 TCC 543  
 RESULT 5  
 AR118686  
 LOCUS AR118686 5360 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 53 from patent US 6150087.  
 ACCESSION AR118686  
 VERSION AR118686.1 GI:14100596  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 5360)  
 AUTHORS Chien,D.Y.  
 TITLE NANBV diagnostics and vaccines  
 JOURNAL Patent: US 6150087-A 53 21-NOV-2000;  
 FEATURES Location/Qualifiers  
 source  
 1..5360  
 /organism="unknown"  
 BASE COUNT 1060 a 1623 c 1532 g 1145 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,226-67 Length: 5360  
 Score: 943.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 98.35% Mismatches: 0  
 Query Match: 98.95% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-965-594-1 (1-182) x AR118686 (1-5360)  
 Qy 1 MetaLapProIleThrAlaTyAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
 Db 930 CTGCGGCCATCAGCGGTACGCCAGCAGACAAAGGGGCTCTAGGGTGCAATATCACC 989  
 Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
 Db 990 AGCCTAACTGCCGGGACAAACCAAGTGGAGGGTGGTCCAGATTGTGCTCACTGCT 1049  
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60  
 Db 1050 GCCCAACCTCTCTGCGACGTCATCAATGGGGTGTGCTGGACTGTCTACCAAGGGGCT 1109  
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetThrThrAsnValAsp 80  
 Db 1110 GGAACAGAGGACATCGCGTCCACCAAGGGTCTCTGTATCCAGATGTATACCAATGTAGAC 1169  
 Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrThrProCysThrCys 100  
 Db 1170 CAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCGGCTCATTTGACACCCCTGCCTTGC 1229  
 Qy 101 GlySerSerAspLeuThrLeuValThrArgHisAlaAspValIleProValArgArgArg 120  
 Db 1230 GGCTCTCGGACCTTTACCTGGTGCAGGACGCGCGATGTCTATCCCTGCGCGGGGCT 1289  
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
 Db 1290 GGTGATAGAGGGGACGCTGTCTGCGCGCGGCCCATTTCTACTTGAAGGCTCCTCGG 1349  
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
 Db 1350 GGGGGTCCCTGTGTGTCGCCCGGGGCGACCGCTGGGCATATTTAGGGCGCGGTGTGC 1409

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrIleThrMet 180  
Db 1410 ACCCGTGGAGTGGCTAAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469  
QY 181 ArgSer 182  
Db 1470 AGGTCC 1475  
RESULT 7  
LOCUS I09328  
DEFINITION Sequence 8 from Patent WO 8904669. DNA linear PAT 02-DEC-1994  
ACCESSION I09328  
VERSION I09328.1 GI:587963  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;  
FEATURES  
Location/Qualifiers  
source  
1..5360  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.22e-67 Length: 5360  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
Gaps: 0  
DB: 6  
US-09-965-594-1 (1-182) x I06434 (1-5360)  
QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
Db 930 CTGGCGCCCATCAGCGGTACGCCAGCAGACAAAGGGCCCTCTAGGGTGCATAATCACC 989  
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 990 AGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCAACTGCT 1049  
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
Db 1050 GCCCAAACTTCCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGGGCC 1109  
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
Db 1110 GGAACGAGGACCATCGCGTACCAAGGGTCTGTCTCATCATGATGTATACCAATGTAGAC 1169  
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1170 CAAGACCTTGTGGGTGGCGCGCTCCGCCAAGGTAGCGCTCATTGACACCTGCATTCG 1229  
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
Db 1230 GGGTCTCGGACCTTACCTGGTTCACGAGGACGCGCATATTTCCTCCGCGCGCGG 1289  
QY 121 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
Db 1290 GGTGATAGAGGGGACGCGCTGTCTGCCCGCCCATTTCTTACTTGAAGGCTCCTCG 1349  
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
Db 1290 GGTGATAGAGGGGACGCGCTGTCTGCCCGCCCATTTCTTACTTGAAGGCTCCTCG 1349  
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValIleGlyIlePheArgAlaValCys 160  
Db 1350 GGGGGTCCGCTGTCTGGCGGGGACGCGGTGGGCATATTATAGGCGCGGTGTGC 1409  
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrIleThrMet 180  
Db 1410 ACCCGTGGAGTGGCTAAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469  
QY 181 ArgSer 182  
Db 1470 AGGTCC 1475  
RESULT 6  
LOCUS I06434  
DEFINITION Sequence 48 from Patent EP 0318216. DNA linear PAT 02-DEC-1994  
ACCESSION I06434  
VERSION I06434.1 GI:590311  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: EP 0318216-A1 48 31-MAY-1989;  
FEATURES  
Location/Qualifiers  
source  
1..5360  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.22e-67 Length: 5360  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
Gaps: 0  
DB: 6  
US-09-965-594-1 (1-182) x I06434 (1-5360)  
QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
Db 930 CTGGCGCCCATCAGCGGTACGCCAGCAGACAAAGGGCCCTCTAGGGTGCATAATCACC 989  
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 990 AGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCAACTGCT 1049  
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
Db 1050 GCCCAAACTTCCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGGGCC 1109  
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
Db 1110 GGAACGAGGACCATCGCGTACCAAGGGTCTGTCTCATCATGATGTATACCAATGTAGAC 1169  
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1170 CAAGACCTTGTGGGTGGCGCGCTCCGCCAAGGTAGCGCTCATTGACACCTGCATTCG 1229  
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
Db 1290 GGTGATAGAGGGGACGCGCTGTCTGCCCGCCCATTTCTTACTTGAAGGCTCCTCG 1349  
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
Db 1290 GGTGATAGAGGGGACGCGCTGTCTGCCCGCCCATTTCTTACTTGAAGGCTCCTCG 1349  
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValIleGlyIlePheArgAlaValCys 160  
Db 1350 GGGGGTCCGCTGTCTGGCGGGGACGCGGTGGGCATATTATAGGCGCGGTGTGC 1409  
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrIleThrMet 180  
Db 1410 ACCCGTGGAGTGGCTAAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469  
QY 181 ArgSer 182  
Db 1470 AGGTCC 1475

Db 1410 ACCCGTGGAGTGGCTAAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469  
QY 181 ArgSer 182  
Db 1470 AGGTCC 1475  
RESULT 7  
LOCUS I09328  
DEFINITION Sequence 8 from Patent WO 8904669. DNA linear PAT 02-DEC-1994  
ACCESSION I09328  
VERSION I09328.1 GI:587963  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;  
FEATURES  
Location/Qualifiers  
source  
1..5360  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.22e-67 Length: 5360  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
Gaps: 0  
DB: 6  
US-09-965-594-1 (1-182) x I09328 (1-5360)  
QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
Db 930 CTGGCGCCCATCAGCGGTACGCCAGCAGACAAAGGGCCCTCTAGGGTGCATAATCACC 989  
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 990 AGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCAACTGCT 1049  
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
Db 1050 GCCCAAACTTCCTGGCAACGTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGGGCC 1109  
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
Db 1110 GGAACGAGGACCATCGCGTACCAAGGGTCTGTCTCATCATGATGTATACCAATGTAGAC 1169  
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1170 CAAGACCTTGTGGGTGGCGCGCTCCGCCAAGGTAGCGCTCATTGACACCTGCATTCG 1229  
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
Db 1230 GGGTCTCGGACCTTACCTGGTTCACGAGGACGCGCATATTTCCTCCGCGCGCGG 1289  
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
Db 1290 GGTGATAGAGGGGACGCGCTGTCTGCCCGCCCATTTCTTACTTGAAGGCTCCTCG 1349  
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValIleGlyIlePheArgAlaValCys 160  
Db 1350 GGGGGTCCGCTGTCTGGCGGGGACGCGGTGGGCATATTATAGGCGCGGTGTGC 1409  
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrIleThrMet 180  
Db 1410 ACCCGTGGAGTGGCTAAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAACCATG 1469  
QY 181 ArgSer 182  
Db 1470 AGGTCC 1475

Db 1470 AGGTCC 1475

RESULT 8  
AR118692  
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 65 from patent US 6150087.  
ACCESSION AR118692  
VERSION AR118692.1 GI:14100602  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 6785)  
AUTHORS Chien,D.Y.  
TITLE NANEV diagnostics and vaccines  
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;  
FEATURES  
Location/Qualifiers  
1..6785  
/organism="unknown"  
BASE COUNT 1392 a 2050 c 1914 g 1429 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.15e-67 Length: 6785  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118692 (1-6785)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
:::|||||  
Db 1203 CTGGCGCCCATCACGGGTACGCCAGCAGACAAAGGGGCTCTAGGGTGCATAATCACC 1262

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
|||  
Db 1263 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATGTGTCAACTGCT 1322

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
|||  
Db 1323 GCCCAACCTTCTTGGCAACGTCATCAATGGGTGTGCTGAGCTGTCTACACGGGGCC 1382

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
|||  
Db 1383 GGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 1442

QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
|||  
Db 1443 CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCGGCTCATTCACACCTGCACCTGC 1502

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
|||  
Db 1503 GGTCTCTGGACCTTTACTGTCTCAGGAGGACGCCGATGTCATTCCTGCGCGCGCGG 1562

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
|||  
Db 1563 GGTGATAGCAGGGCAGGCTCTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1622

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
|||  
Db 1623 GGGGTCTGGCTGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1682

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
|||  
Db 1683 ACCCTGGAGTGGCTAAGCGGTGACCTTTATCCCTGTGGAGAACCTAGAGACCAACCATG 1742

QY 181 ArgSer 182  
|||||  
Db 1743 AGGTCC 1748

RESULT 9  
Db 1743 AGGTCC 1748

I06440  
LOCUS I06440 6785 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 54 from Patent EP 0318216.  
ACCESSION I06440  
VERSION I06440.1 GI:590312  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 6785)  
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: EP 0318216-A1 54 31-MAY-1989;  
FEATURES  
Location/Qualifiers  
1..6785  
/organism="unknown"  
BASE COUNT 1392 a 2050 c 1914 g 1429 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.15e-67 Length: 6785  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x I06440 (1-6785)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
:::|||||  
Db 1203 CTGGCGCCCATCACGGGTACGCCAGCAGACAAAGGGGCTCTAGGGTGCATAATCACC 1262

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
|||  
Db 1263 AGCCTAACTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATGTGTCAACTGCT 1322

QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
|||  
Db 1323 GCCCAACCTTCTTGGCAACGTCATCAATGGGTGTGCTGAGCTGTCTACACGGGGCC 1382

QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
|||  
Db 1383 GGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTAGAC 1442

QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
|||  
Db 1443 CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCGGCTCATTCACACCTGCACCTGC 1502

QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
|||  
Db 1503 GGTCTCTGGACCTTTACTGTCTCAGGAGGACGCCGATGTCATTCCTGCGCGCGCGG 1562

QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
|||  
Db 1563 GGTGATAGCAGGGCAGGCTCTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1622

QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
|||  
Db 1623 GGGGTCTGGCTGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1682

QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
|||  
Db 1683 ACCCTGGAGTGGCTAAGCGGTGACCTTTATCCCTGTGGAGAACCTAGAGACCAACCATG 1742

QY 181 ArgSer 182  
|||||  
Db 1743 AGGTCC 1748

RESULT 10  
I09329  
LOCUS I09329 6785 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 10 from Patent WO 8904669.

ACCESSION I09329  
VERSION I09329.1 GI:587964  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 6785)  
AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.  
JOURNAL Patent: WO 8904669-A 10 01-JUN-1989;  
FEATURES Location/Qualifiers  
source 1..6785  
BASE COUNT 1392 a 2050 c 1914 g 1429 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,150-67 Length: 6785  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservatative: 3  
Best Local Similarity: 98.35% Mismatches: 3  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0  
US-09-965-594-1 (1-182) x I09329 (1-6785)  
Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20  
Db 1203 CTGGCCCATCAGCGGTACGCCAGCAGACAGGGGCCCTCTAGGGTGCATATCACC 1262  
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 1263 AGCTTAACCTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 1322  
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60  
Db 1323 GCCCAACCTTCTCGCAACGTGCATCAATGGGTGTGCTGACTGTCTACCCAGGGGCC 1382  
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnValAsp 80  
Db 1383 GGAACGAGGACCATCGCTCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC 1442  
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1443 CAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTTGCATTGC 1502  
Qy 101 GlySerSerAspLeuTyrIleValThrArgHisAlaAspValIleProValArgArg 120  
Db 1503 GGCTCTCGACCTTTACCTGTGTCAGGAGCAGCGCGATGTATTCCCGTGGCGGGCGG 1562  
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrIleLysGlySerSer 140  
Db 1563 GGTGATAGCAGGGGACGCTGTGTCGCCCGGCCCAATTTCTTACTTGAAGGCTCCTCG 1622  
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
Db 1623 GGGGGTCCGCTGTGTGCCCGGGGCGAGCGCGTGGGCATATTTAGGCGCGGGTGTGC 1682  
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
Db 1683 ACCCGTGGACTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 1742  
Qy 181 ArgSer 182  
Db 1743 AGGTCC 1748  
RESULT 11  
LOCUS AR118696 7310 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 74 from patent US 6150087.  
ACCESSION AR118696  
VERSION AR118696.1 GI:14100606  
KEYWORDS  
SOURCE Unknown.

ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7310)  
AUTHORS Chien, D.Y.  
JOURNAL Patent: US 6150087-A 74 21-NOV-2000;  
FEATURES Location/Qualifiers  
source 1..7310  
BASE COUNT 1495 a 2220 c 2056 g 1539 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,5e-67 Length: 7310  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservatative: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0  
US-09-965-594-1 (1-182) x AR118696 (1-7310)  
Qy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuGlyCysIleIleThr 20  
Db 1728 CTGGCCCATCAGCGGTACGCCAGCAGACAGGGGCCCTCTAGGGTGCATATCACC 1787  
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 1788 AGCTTAACCTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 1847  
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60  
Db 1848 GCCCAACCTTCTCGCAACGTGCATCAATGGGTGTGCTGACTGTCTACCCAGGGGCC 1907  
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnValAsp 80  
Db 1908 GGAACGAGGACCATCGCTCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC 1967  
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1968 CAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCTTGCATTGC 2027  
Qy 101 GlySerSerAspLeuTyrIleValThrArgHisAlaAspValIleProValArgArg 120  
Db 2028 GGCTCTCGACCTTTACCTGTGTCAGGAGCAGCGCGATGTATTCCCGTGGCGGGCGG 2087  
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrIleLysGlySerSer 140  
Db 2088 GGTGATAGCAGGGGACGCTGTGTCGCCCGGCCCAATTTCTTACTTGAAGGCTCCTCG 2147  
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
Db 2148 GGGGGTCCGCTGTGTGCCCGGGGCGAGCGCGTGGGCATATTTAGGCGCGGGTGTGC 2207  
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
Db 2208 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACCATG 2267  
Qy 181 ArgSer 182  
Db 2268 AGGTCC 2273  
RESULT 12  
LOCUS I09331 7310 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 15 from Patent WO 8904669.  
ACCESSION I09331  
VERSION I09331.1 GI:587966  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7310)

AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.  
JOURNAL Patent: WO 8904669-A 15 01-JUN-1989;  
FEATURES Location/Qualifiers

Source 1..7310  
BASE COUNT 1495 a 2218 c 2058 g 1539 t  
ORIGIN /organism="unknown"

Alignment Scores:  
Pred. No.: 4,5e-67 Length: 7310  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservativeness: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x 109331 (1-7310)

QY 1 MetaAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
Db 1728 CTGGCGCCCATCACGGCTACGCCAGCAGACAGGGCCCTCTAGGTCATCAATACACC 1787  
QY 21 SerLeuThrGlyArgAspIleAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
Db 1788 AGCCTAAGTGGCGGGGAGAAACCAAGTGGAGGGTGAAGTCCAGATTGTCAACTGCT 1847  
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60  
Db 1848 GCCCAACCTTCTCGCAACGTGCATCAATCGGTGTCTGGACTGTCTACCGAGGGCC 1907  
QY 61 GlyThrArgThrIleAlaSerProIleGlyProValIleGlnMetTyrThrAsnValAsp 80  
Db 1908 GGAACGAGGACCATCGCTCACCAAGGCTCTGTCAATCCAGATGTATACCAATGTAGAC 1967  
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
Db 1968 CAAGACCTTGTGGCTGCGCCGCTCCGCAAGTAGAGCGCTCATTTGACACCTGCCTGC 2027  
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120  
Db 2028 GGCCTCTGGACCTTACTGTGTGCGCGCGGCGACCGCATTTTCCGCTGCGCGCGCGCG 2087  
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
Db 2088 GTGTATACAGGGGAGCGCTGCTGCGCGCGCGCCCTTCTACTTGAAGGCTCTCTCG 2147  
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
Db 2148 GGGGCTCGCTGTTGTGCGCGCGGCGGCGACCGCTGGGCATATTATGGCGCGGTGTC 2207  
QY 161 ThrArgGlyValAlaValAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
Db 2208 ACCCTGTAGTGGCTTGAAGCGGCGGAGCTTTAATCCCTGTGGAGAACCTAGACACACCT 2267  
QY 181 ArgSer 182  
Db 2268 AGGTCC 2273

RESULT 13  
LOCUS HPCPOLYP  
DEFINITION Hepatitis C virus polyprotein gene, partial cds.  
ACCESSION M32084  
VERSION M32084.1 G1:329875  
KEYWORDS polyprotein.  
SOURCE Hepatitis C virus  
ORGANISM Hepatitis C virus  
Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;  
Hepadnaviruses.  
REFERENCE 1 (bases 1 to 7310)  
AUTHORS Choo, Q.-K., Richman, K. and Han, J.  
TITLE The nucleotide sequence of the Hepatitis C viral genome  
JOURNAL Unpublished (1990)

COMMENT

Original source text: Hepatitis C virus, cDNA to viral RNA, clones K9-1 through 15e, isolated from chimpanzee (individual 910) blood plasma.

Draft entry and printed sequence for [1] kindly submitted by M.Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608.

FEATURES

source

Location/Qualifiers  
1..7310  
/organism="Hepatitis C virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:11103"  
<1..>7310  
/note="polyprotein"  
/codon\_start=3  
/protein\_id="AAA45677.1"  
/db\_xref="GI:329876"  
/translation="GCPERLASCKPLTFDQGMCPISYANGSGPDRQPCWHYPPKPC  
GIVPAKSCGVPYCTPDSVVTGTSWENDTDVFLNTPRPLGNGFC  
TWNSTGTVKCGAPCVIGGAGNTLHCTDFCRKHPDATYSCGSDPMTPLCLVD  
YPRLMHYPTCTINVTIFKIMYGVGVEHLEACNTRGCRCDLERDSELSPLLT  
TDQVLPCTSTLPLASTGLIHLHONIVDYLYXGSSIASHAKIKWEYVLLFLL  
ADARVCSCLMHLLISOREALENLYILNASSIAGTHLVFLVFFCFAYLGNKWP  
GAVTTFYGMPLLLLLALPORAYALDEVAASCGVVLVGLMALTSYKRYISWC  
LWMLQYILTRVFAQLHWIPLPNVGRGRDAILIMCAVHTLVFDIKLLAVFGLW  
ILOASLLKVPYFVRVGLLRFCALARKMIGHYVQWYIKLGALTCTYVNLPLRD  
WAHNGRLDLAVAVPEVFSOMETKLITWAGDTAACGDIINGLPSARRRRIILGPAD  
GMVSGRLLAPITAYAQGTGLLGCITSLTGRDKNOVEGVQIVSTRAQTFLATCI  
NGVQWYVHCAGTFTIAPKGPVLOMYNDQDLVGMPPAQSGRSITPTCCASSDYL  
NTRADIVPVRGDSRGSLLSPRISYLAGSGGGLPCPAGHAGVIFRAAYATGVA  
KAVDFIPVENLETTMSPVTDNSPVPQSFQVAHLHAPTGSKSTKVPAAVAAQV  
GYLVLPASVNAATLFGAYSKAHGIDPNITGRTITTTGSPITSTYKGLFADGCS  
KKATIIICDECHSDATSLIGTDLQDAETAGARLVATATPPGTVTPHPNTEE  
VALSTGEIPYKNAIPLVYIKGGRHLIFCHSKKQDELAALVALGINAVYRGLD  
VSVPTSGDVVATDALMTGYTGDPSVIDCNTCTVTDVDFSLDPTFTIETLPQD  
AVSTRRGRTGRKPGIYRFVAPGRPSCHFDSSVLCQYDAGCAWYELTPTFTVR  
LRYMNTPGLVPCODHLEFEGVFTGLTHDAHFLSOTKSGENLPLVAYQATVCA  
AQRPPSDQMMKCLIRKLTFLHGTPLDLRLGAYQNEITLTHPTKTIIMTMSADLE  
VYTSVLVGGVLAALAAAYCLSTGVVIVGVLSGRPALIPDRELYREFDEMECS  
QHLPTIYOGMMLEQFKALGLLQATASQARVIAPAVQTNMOKLETFWAKHMFNIS  
GIQYLAGSLTLPNPATIASMAETAAVTSPLTTSQTLLENILGGWAAQALAPGANA  
FVAGLAGAAIGSVGLKGLVIDILAGYAGVAGALVAFKIMSGEVPSTBDLVNLLPAI  
LSPALVGVVCAAILRRHVGCGEAVQVMNRLIAFASGNHVSPTHYVPSDAARV  
TALLSSLTOLLRLHQLWISSECTTCSGSLRDIDWICEVLSDFKTLAKLMQ  
LPFGVSCQYKGVWVDGIMHTRCHGAEITGVKNGTMRIVGPRTRCNMNSGTF  
PINATYDGPCLPAPNTFALMRVSAEYVEIRQVDGFHYVTGMTDNLKPCQVYS  
PEFTLEDGLVRLHRAFPCKPLREYFVGVHLEYVPGSQLCEPPEPDAVTSMT  
DPSHITFAAGRLARGSPSSASSASQSLKATCANHSDPSDAELIEANILWR  
OMGNTITRVESKNVILSDPLVAEEDEREISVPAEILKRSRFAOALPVWARP  
YNPLVETKKPDPVPPVHCGCLPPKPPSPVPPPKRTVLTSTLSTALAEALTR  
SFGSSSTSGITGDNATTSSEPAEGCPSDASEYSMPLEGEPEGDPDLSGNSY  
SSEANAEDVCCSISYWTGALVTPCAAEQKLPINALSNLSLRLHNLVSTTSRAC  
QRKRVTFDRQLVDLSDHQDLREVKAASKVKANLUSVEEACSLTPPHSAKSKFYG  
AKDVRHARKAVTHINSVKDLLEDNVTPIIDTTIMAKNEVFCVQPKGRKPARLIV  
PDLGVYCEKMAVDVTKPLPVMGSSYFQYSPQORVEFLVQAKSKKTPMGFSYD  
TRCFDSTVIESDIETEEIYOCDDLPQARVAIKSLTERLYVGGLTNSRGENGVRR  
CRASGLTSCGNTLTCYIKARAACRAAGLQDCTMLVCGDDLVLYCESAGVQDAASL  
RAFTAMTRYSAAPGDPPEYDLELITSCSSNVVAHDCAGKRVYLTPTPTPLAR  
AAWETARHTPVNSHLNLIHFAPTLHARNILMTHTFSVLIARDQALDCEIYGACY  
SIEPLDLPPIIQLR"

BASE COUNT 1495 a 2218 c 2058 g 1539 t  
ORIGIN

Alignment Scores:

Pred. No.: 4,5e-67 Length: 7310  
Score: 943.00 Matches: 179  
Percent Similarity: 100.00% Conservativeness: 3  
Best Local Similarity: 98.35% Mismatches: 0  
Query Match: 98.95% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x HPCPOLYP (1-7310)

QY 1 MetaAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20

```

:::
1728 CTGGCGCCATACGGCGTACGCCACACAGGGCGCTCTAGGGTCATAATCACC 1787
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db AGCCTAACTGCGCGGACAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853
1788 AGCCTAACTGCGCGGACAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1847
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db GCCAAACCTTCTCGGCAACGTGCATCAATGGGGTGCTGGACTGTCTACCAACGGGGCC 2913
61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
1908 GGAAGGAGACCATCGCGTCACCAAGGTCCTGCATCCAGATGTATACCAATGTAGAC 2973
QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCAITGACACACCTGCACCTGC 2027
101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
2028 GGCTCTCGGACCTTACCTGTCACAGGACGCGCAITGTCATCCAGATGTATACCAATGTAGAC 1967
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db GGTGATACAGGGGACGCTGCTGCGCGCGGCCATTTCTACTTGAAGGCTCCTCG 3093
2088 GGTGATACAGGGGACGCTGCTGCGCGCGGCCATTTCTACTTGAAGGCTCCTCG 2147
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValCys 160
Db GGGGTCTCGCTGTGTGCGCGCGGACCGCTGGGCATATTAGGGCGCGGTGTC 2267
161 ThrArgGlyValAlaAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
2208 ACCGTGGAGTGGCTAAGCGGTGGACTTTATCCTGTGGAGAACCTAGAGACAACCATG 2267
QY 181 ArgSer 182
Db AGGTCC 2273

RESULT 14
AR118703
LOCUS AR118703 8316 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 88 from patent US 6150087.
ACCESSION AR118703
VERSION AR118703.1 GI:14100613
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8316)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;
FEATURES
    source
        1..8316
        /organism="unknown"
BASE COUNT 1671 a 2529 c 2345 g 1771 t
ORIGIN

Alignment Scores:
Pred. No.: 5,17e-67 Length: 8316
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118703 (1-8316)

QY 1 MetaLaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuCysIleIleThr 20
Db CTGGCGCCCATCAGCGCGTACGCCACAGACAGAGGGGCGCTCTAGGTGTCATAATCACC 2793
```

```

QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db AGCCTAACTGCGCGGACAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 2853
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db GCCAAACCTTCTCGGCAACGTGCATCAATGGGGTGCTGGACTGTCTACCAACGGGGCC 2913
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db GGAAGGAGACCATCGCGTCACCAAGGTCCTGCATCCAGATGTATACCAATGTAGAC 2973
QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db CAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCAITGACACACCTGCACCTGC 3093
2974 CANGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCAITGACACACCTGCACCTGC 3033
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
Db GGTGATACAGGGGACGCTGCTGCGCGCGGCCATTTCTACTTGAAGGCTCCTCG 3153
3034 GGTGATACAGGGGACGCTGCTGCGCGCGGCCATTTCTACTTGAAGGCTCCTCG 3093
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db GGTGATACAGGGGACGCTGCTGCGCGCGGCCATTTCTACTTGAAGGCTCCTCG 3153
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValCys 160
Db GGGGTCTCGCTGTGTGCGCGCGGACCGCTGGGCATATTAGGGCGCGGTGTC 3213
161 ThrArgGlyValAlaAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
3214 ACCGTGGAGTGGCTAAGCGGTGGACTTTATCCTGTGGAGAACCTAGAGACAACCATG 3273
QY 181 ArgSer 182
Db AGGTCC 3279

RESULT 15
AR118728
LOCUS AR118728 8987 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 137 from patent US 6150087.
ACCESSION AR118728
VERSION AR118728.1 GI:14100638
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8987)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 137 21-NOV-2000;
FEATURES
    source
        1..8987
        /organism="unknown"
BASE COUNT 1807 a 2735 c 2547 g 1898 t
ORIGIN

Alignment Scores:
Pred. No.: 5,62e-67 Length: 8987
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 6 Gaps: 0

US-09-965-594-1 (1-182) x AR118728 (1-8987)

QY 1 MetaLaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuCysIleIleThr 20
Db CTGGCGCCCATCAGCGCGTACGCCACAGACAGAGGGGCGCTCTAGGTGTCATAATCACC 3135
21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db AGCCTAACTGCGCGGACAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 3195
```



```
Qy      41  AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db      3196  GCCCAAAACCTTCCTGGCAAGTCGATCAATGGGGTGTGTGGACTGTCTACCAAGGGGCC 3255
Qy
Db      61  GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db      3256  GGAACGAGGACCATCGGTCACCCCAAGGGTCTGTGTCATCCAGATGTATACCAATGTAGAC 3315
Qy      81  LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerIleuThrProCysThrCys 100
Db      3316  CAAGACCTTGTGGCTGGCCGCTCGCAAGGTAGCCGCTCATTTGACACCTGCACCTTGC 3375
Qy      101  GlySerSerAspLeuTyrIleuValThrArgHisAlaAspValIleProValArgArgArg 120
Db      3376  GGCTCCTCGGACCTTTACCTGGTCAGAGCGCAGCCGATGTCATCCCGTGGCCGGCGG 3435
Qy      121  GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrIleuLysGlySerSer 140
Db      3436  GGTGATAGCAGGGGCGAGCCTGCTGTGCCCGGCCCATTTCTACTTGAAGGCTCCTCG 3495
Qy      141  GlyGlyProLeuLeuCysProAlaGlyHisAlaValAlcIlyIlePheArgAlaAlaValCys 160
Db      3496  GGGGTCCTGCTGTGTGCCCGGGGCGACCGCGTGGGCATATTAGGGCCGGGTGTGC 3555
Qy      161  ThrArgGlyValAlaLysAlaValAspPheIleProValIleuSerLeuGluThrThrMet 180
Db      3556  ACCCGTGGAGTGGCTAAGCGCGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 3615
Qy      181  ArgSer 182
Db      3616  AGGTCC 3621
```

Search completed: August 31, 2003, 00:45:16  
Job time : 2382.6 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 41.2251 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQTRGLGCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/gcneseq/gcneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	953	100.0	182	21	AA15211 Hepatitis C virus
2	946	99.3	686	23	AA18689 HCV-1 NS3/4a mutan
3	946	99.3	686	23	AAU76377 Hepatitis C virus
4	946	99.3	686	24	ABG72261 HCV-1 NS3/4a confo
5	944	99.1	3011	14	AA40120 HCV genomic amino
6	943	99.0	609	15	AA451170 Hepatitis C virus
7	943	99.0	1766	10	AA192041 Sequence encoded i
8	943	99.0	1786	10	AA190158 Protein sequence o
9	943	99.0	2261	10	AA190164 Peptide encoded by

10	943	99.0	2301	10	AA192047 Sequence encoded i
11	943	99.0	2436	10	AA192050 Sequence encoded i
12	943	99.0	2436	10	AA192088 Peptide encoded by
13	943	99.0	2772	21	AA18540 Protein encoded by
14	943	99.0	2816	14	AA18540 HCV-1 polyprotein.
15	943	99.0	2894	16	AA170230 Composite hepatiti
16	943	99.0	2955	20	AA14975 Amino acid sequenc
17	943	99.0	2955	21	AA18541 Polyprotein encode
18	943	99.0	3011	13	AA18541 Compiled HCV seque
19	943	99.0	3011	14	AA18541 Hepatitis C virus
20	943	99.0	3011	17	AA18541 Hepatitis C virus
21	943	99.0	3011	18	AA18541 HCV polyprotein.
22	943	99.0	3011	19	AA18541 HCV polyprotein.
23	943	99.0	3011	23	AA18541 Hepatitis C virus
24	943	99.0	3011	23	AA18541 Hepatitis C virus
25	942	98.8	632	23	AA18541 Hepatitis C virus
26	942	98.8	632	23	AA18541 Hepatitis C virus
27	942	98.8	686	23	AA18541 Hepatitis C virus
28	942	98.8	686	23	AA18541 Hepatitis C virus
29	942	98.8	686	23	AA18541 Hepatitis C virus
30	942	98.8	686	23	AA18541 Hepatitis C virus
31	942	98.8	686	23	AA18541 Hepatitis C virus
32	942	98.8	686	23	AA18541 Hepatitis C virus
33	942	98.8	686	23	AA18541 Hepatitis C virus
34	942	98.8	686	23	AA18541 Hepatitis C virus
35	942	98.8	686	23	AA18541 Hepatitis C virus
36	942	98.8	686	23	AA18541 Hepatitis C virus
37	942	98.8	686	23	AA18541 Hepatitis C virus
38	942	98.8	686	23	AA18541 Hepatitis C virus
39	942	98.8	686	23	AA18541 Hepatitis C virus
40	942	98.8	686	23	AA18541 Hepatitis C virus
41	942	98.8	686	23	AA18541 Hepatitis C virus
42	942	98.8	686	23	AA18541 Hepatitis C virus
43	942	98.8	686	23	AA18541 Hepatitis C virus
44	942	98.8	686	23	AA18541 Hepatitis C virus
45	942	98.8	686	23	AA18541 Hepatitis C virus

## ALIGNMENTS

RESULT 1  
ID AAB15211 standard; protein; 182 AA.

AC AAB15211;

XX 19-DEC-2000 (first entry)

DE Hepatitis C virus NS3 protease.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer.

XX Hepatitis C virus.

XX WO200040707-A1.

PN 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX N-PSDB; AAA70344.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1 substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT

XX

PS

XX

XX

Claim 3; Fig 9; 66pp; English.

XX

CC The present sequence is the Hepatitis C virus (HCV) NS3 protease enzyme.

CC This protein is essential for the replication of the virus, acting to

CC cleave its replicative proteins from the polypeptide produced from the

CC HCV genome. NS4A is also needed for this process and inhibitors of the

CC two proteins should act as antiviral treatments of HCV infection. This is

CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver

CC failure and liver cancer. The present invention concerns a number of NS3

CC mutants and NS3-NS4A fusion proteins which can be used to identify

CC inhibitors of this type as well as enabling structural studies of the

CC protease and protease-inhibitor complexes.

XX

SQ Sequence 182 AA;

Query Match 100.0%; Score 953; DB 21; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.9e-91;

Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGA 60

DB 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGA 60

QY 61 GTRTIASPKGPVIOYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

DB 61 GTRTIASPKGPVIOYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

QY 121 GDSRGSLLSPRISYILKSGSGGLPCPAGHVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

DB 121 GDSRGSLLSPRISYILKSGSGGLPCPAGHVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

QY 181 RS 182

DB 181 RS 182

RESULT 2

AAE18689

ID AAE18689 standard; Protein; 686 AA.

XX

AC AAE18689;

XX

DT 17-MAY-2002 (first entry)

XX

XX HCV-1 NS3/4a mutant conformational antigen.

XX

XX Hepatitis C virus; NS3/4a antigen; HCV infection; mutant; mutein.

XX

OS Hepatitis C virus type 1.

OS Synthetic.

XX

XX Key Location/Qualifiers

FT Misc-difference 403

FT Misc-difference /note= "Wild type Thr substituted with Pro"

FT Misc-difference 404

FT Misc-difference /note= "Wild type Ser substituted with Ile"

XX

XX WO200196875-A2.

XX

XX 20-DEC-2001.

XX

XX 14-JUN-2001; 2001WO-US19369.

XX

XX 15-JUN-2000; 2000US-212082P.

XX

XX 02-APR-2001; 2001US-280811P.

XX

XX 02-APR-2001; 2001US-280867P.

XX

XX (CHIR ) CHIRON CORP.

XX

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

PI

PI Medina-Selby A;

XX

DR WP1; 2002-179522/23.

DR N-PSDB; AAD29795.

XX

XX Immunoassay solid support useful for detecting hepatitis C virus

PT infection in a biological sample, comprises at least one of HCV

PT anti-core antibody and HCV NS3/4a epitope, bound to the support

XX

XX Example 2; Fig 4; 87pp; English.

PS

CC The present invention relates to hepatitis C virus (HCV) core antigen

CC and NS (nonstructural) 3/4a antibody combination assay that can detect

CC both HCV antigens and antibodies present in a sample using a single

CC solid matrix as well as immunoassay solid supports for use in the assay.

CC The solid support is useful for detecting HCV infection in a biological

CC sample. The present sequence is HCV-1 NS3/4a mutant conformational

CC antigen. This sequence is used in the exemplification of the invention.

XX

SQ Sequence 686 AA;

Query Match 99.3%; Score 946; DB 23; Length 686;

Best Local Similarity 98.9%; Pred. NO. 1.2e-89;

Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGA 60

DB 1 MAPITAYAOQTGRLGCIITSLTGRDKNQVEGEVIVSTAAQTFLATCINGVCWTVYHGA 60

QY 61 GTRTIASPKGPVIOYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

DB 61 GTRTIASPKGPVIOYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

QY 121 GDSRGSLLSPRISYILKSGSGGLPCPAGHVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

DB 121 GDSRGSLLSPRISYILKSGSGGLPCPAGHVGIFRAAVCTRGVAKAVDFIPVESLETTM 180

QY 181 RS 182

DB 181 RS 182

RESULT 3

AAU76377

ID AAU76377 standard; Protein; 686 AA.

XX

AC AAU76377;

XX

DT 08-MAY-2002 (first entry)

XX

XX Hepatitis C virus NS3/4a conformational epitope protein sequence.

XX

XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;

XX

XX immunoassay solid support; multiple epitope fusion antigen; MEFA;

XX

XX non-structural protein; mutant; mutein.

XX

OS Hepatitis C virus.

OS Synthetic.

XX

XX Key Location/Qualifiers

FT Misc-difference 403

FT Misc-difference /note= "Wild-type Thr substituted by Pro"

FT Misc-difference 404

FT Misc-difference /note= "Wild-type Ser substituted by Ile"

XX

XX WO200196870-A2.

XX

XX 20-DEC-2001.

XX

XX 14-JUN-2001; 2001WO-US19156.

XX

XX 15-JUN-2000; 2000US-212082P.

XX

XX 02-APR-2001; 2001US-280811P.

XX

XX

```

PR 02-APR-2001; 2001US-280867P.
XX (CHIR ) CHIRON CORP.
PA Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
XX WPI: 2002-090228/12.
DR N-PSDB; ABK15344.
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support .
XX Claim 5; Fig 3; 92pp: English.
XX The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the non-structural protein NS3/4a conformational epitope of
CC the invention.
XX Sequence 586 AA:
XX
XX Query Match 99.3%; Score 946; DB 23; Length 686;
XX Best Local Similarity 98.9%; Pred. No. 1.2e-89;
XX Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0:
QY 1 MAPITAYAOOTRGLGCIITSLTGRDNKQVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 60
DB 1 MAPITAYAOOTRGLGCIITSLTGRDNKQVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 60
QY 61 GTRTIASPKGPVIQMYTNYVDKDLVGMWPAQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120
DB 61 GTRTIASPKGPVIQMYTNYVDQDLVGMWPAQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120
QY 121 GDSRGLSLSPRISYIKGSSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVSLSTTM 180
DB 121 GDSRGLSLSPRISYIKGSSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVLENLTTM 180
QY 181 RS 182
DB 181 RS 182

```

## RESULT 4

ABG72261  
ID ABG72261 standard; Protein; 686 AA.

```

XX AC ABG72261;
XX AC
XX DT 06-MAR-2003 (first entry)
XX DE HCV-1 NS3/4a conformational antigen.
XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
XX NS3/4a conformational epitope; multiple epitope fusion antigen;
KW

```

```

KW MEFA: anti-HCV antibody; NS3/4a conformational antigen;
KW HCV infection; mutant; mutein.
XX
XX Hepatitis C virus type 1.
XX Synthetic.
XX Key Location/Qualifiers
XX Region 2..686
XX Note- "Corresponds to amino acid residues 1027-1711
XX of HCV-1 NS3/4a polypeptide"
XX
XX Misc-difference 403
XX Note- "Substitution of wild-type Thr to Pro"
XX Misc-difference 404
XX Note- "Substitution of wild-type Ser to Ile"
XX
XX US2002146685-A1.
XX
XX 10-OCT-2002.
XX
XX 14-JUN-2001; 2001US-0881654.
XX
XX 15-JUN-2000; 2000US-212082P.
XX 02-APR-2001; 2001US-280811P.
XX 02-APR-2001; 2001US-280867P.
XX
XX (CHIE/) CHIEN D Y.
XX (ARCA/) ARCANGEL P.
XX (TAND/) TANDESKE L.
XX (GEOR/) GEORGE-NASCIMENTO C.
XX (COIT/) COIT D.
XX (MEDI/) MEDINA-SELBY A.
XX
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
XX Medina-Selby A;
XX WPI: 2003-147573/14.
XX N-PSDB; ABX14410.
XX
XX Immunoassay solid support for detecting Hepatitis C Virus infection in
XX biological samples, comprises Hepatitis C Virus conformational epitope
XX and multiple epitope fusion antigen .
XX Claim 2; Fig 3A-3D; 45pp: English.
XX
XX The present invention relates to immunoassays comprising Hepatitis C
XX virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
XX antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
XX the multiple epitope fusion antigen react with anti-HCV antibodies
XX present in a biological sample from an HCV-infected individual. The
XX immunoassays and methods of the invention are useful for detecting
XX HCV infection in a biological sample. The inventive immunoassay solid
XX support provides a sensitive and reliable method for detecting early
XX HCV seroconversion. The assays can detect HCV infection caused by any
XX six known genotypes of HCV. The use of the multiple epitope fusion
XX proteins decreases masking problems, improves sensitivity in detecting
XX antibodies by allowing a greater number of epitopes on a unit area
XX of substrate, and improves selectivity. The present sequence
XX represents HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant
XX of the HCV-1 NS3/4a polypeptide.
XX Sequence 686 AA:
SQ

```

Query Match 99.3%; Score 946; DB 24; Length 686;

Best Local Similarity 98.9%; Pred. No. 1.2e-89;

Matches 180; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTRGLGCIITSLTGRDNKQVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 60

DB 1 MAPITAYAOOTRGLGCIITSLTGRDNKQVEGEVQIVSTAAQTFLATCINGVCMTVYHGA 60

QY 61 GTRTIASPKGPVIQMYTNYVDKDLVGMWPAQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120

DB 61 GTRTIASPKGPVIQMYTNYVDQDLVGMWPAQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120

Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 Db |||||  
 Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
 Db |||||

Qy 181 RS 182  
 ||  
 Db 181 RS 182

RESULT 5  
 AAR40120  
 ID AAR40120 standard; Protein: 3011 AA.  
 XX  
 AC AAR40120;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 27-JAN-1994 (first entry)  
 XX  
 DE HCV genomic amino acid sequence isolated from infected human LG.  
 XX  
 KW Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;  
 KW human growth hormone; HGH; secretion signal; fusion protein;  
 KW vaccine.  
 XX  
 OS Hepatitis C Virus.  
 XX  
 PN WO9315193-A1.  
 XX  
 PD 05-AUG-1993.  
 XX  
 PF 29-JAN-1993; 93WO-US00907.  
 XX  
 PR 31-JAN-1992; 92US-0830024.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Bode SL, Casey JM, Desai SM, Devare SG, Frail DE;  
 PI Yamaguchi J, Zeck BJ;  
 XX  
 DR WPI; 1993-258673/32.  
 XX  
 PT New plasmid pHCV-162 is a mammalian expression systems for HCV  
 PT proteins - useful for diagnosing HCV infection and as vaccines  
 PT for preventing HCV infection  
 XX  
 PS Example 1; Page 39-49; 100pp: English.  
 XX  
 CC RNA was isolated from the plasma of a HCV seropositive human  
 CC (designated "LG") and cDNA was prepared from it. The cDNA was  
 CC PCR amplified using specific primers with sequences based  
 CC on the prototype HCV-1 cDNA sequence (GENBANK M2321). Further  
 CC amplification using nested primers resulted in 7 adjacent HCV DNA  
 CC fragments which could be assembled into a full-length sequence. The  
 CC DNA sequence was determined and translated into the genomic amino  
 CC acid sequence. Comparison of the LG genomic amino acid sequence  
 CC with that from HCV-1 showed 134 amino acid differences.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 3011 AA.

Query Match 99.1%; Score 944; DB 14; Length 3011;  
 Best Local Similarity 98.4%; Pred. No. 1.4e-88;  
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60  
 :|||||  
 Db 1026 LAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085

Qy 61 GTRTIASPKGPVIOYNTVDKLVGHPAPQSGRSITPCTCGSSDLYLVTRHADVIPVRR 120  
 |||||  
 Db 1086 GTRTIASPKGPVIOYNTVDKLVGHPAPQSGRSITPCTCGSSDLYLVTRHADVIPVRR 1145

Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 Db |||||  
 Qy 181 RS 182  
 ||  
 Db 1206 RS 1207

RESULT 6  
 AAR51170  
 ID AAR51170 standard; peptide: 609 AA.  
 XX  
 AC AAR51170;  
 XX  
 DT 20-OCT-1994 (first entry)  
 XX  
 DE Hepatitis C virus non-structural protein 3.  
 XX  
 KW Peptide; antibody; hepatitis C virus; HCV; identification;  
 KW diagnosis; non-A non-B hepatitis; NANB; detection.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PN JP06056891-A.  
 XX  
 PD 01-MAR-1994.  
 XX  
 PF 05-AUG-1992; 92JP-0209201.  
 XX  
 PR 05-AUG-1992; 92JP-0209201.  
 XX  
 PA (OLYU ) OLYMPUS OPTICAL CO LTD.  
 XX  
 DR WPI; 1994-106803/13.  
 XX  
 PI New peptide(s) reactive with anti-hepatitis C virus antibody -  
 PI for specific, early diagnosis of HCV infection  
 XX  
 PS Disclosure; Page 9-10; 15pp; Japanese.  
 XX  
 CC Peptide fragments of the non-structural protein (NS3) are reactive  
 CC with and can detect antibodies against the NS3 domain of HCV. The  
 CC peptides can be used for diagnosis of HCV infection. Nonspecific  
 CC reaction can be inhibited and misdiagnosis of HCV infection can be  
 CC decreased. See AAR51162-70.  
 XX  
 SQ Sequence 609 AA.

Query Match 99.0%; Score 943; DB 15; Length 609;  
 Best Local Similarity 98.4%; Pred. No. 2.1e-89;  
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60  
 :|||||  
 Db 20 LAPITAYAOOTRGLGCIITSITGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTYHGA 79

Qy 61 GTRTIASPKGPVIOYNTVDKLVGHPAPQSGRSITPCTCGSSDLYLVTRHADVIPVRR 120  
 |||||  
 Db 80 GTRTIASPKGPVIOYNTVDKLVGHPAPQSGRSITPCTCGSSDLYLVTRHADVIPVRR 139

Qy 121 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 |||||  
 Db 140 GDSRGLSPRPISYLKSGSGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 199

Qy 181 RS 182  
 ||  
 Db 200 RS 201

RESULT 7  
 AAP92041  
 ID AAP92041 standard; protein: 1765 AA.

```

XX AAP92041;
AC
XX
XX
DT 25-MAR-2003 (updated)
DT 02-MAR-1990 (first entry)
XX
XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f,
DE 33f, 33g and 39c.
XX
XX Hepatitis C virus (HCV): non-A, non-B hepatitis (HANBH)
KW
XX
XX Hepatitis C virus.
OS
XX
XX EP318216-A.
PN
XX
XX 31-MAY-1989.
PD
XX
XX 18-NOV-1988; 88EP-0310522.
PF
XX
XX 18-NOV-1987; 87US-0122714.
PR
XX 30-DEC-1987; 87US-0139886.
PR
XX 26-FEB-1988; 88US-0161072.
PR
XX 06-MAY-1988; 88US-0191263.
PR
XX 26-OCT-1988; 88US-0263584.
PR
XX 14-NOV-1988; 88US-0271450.
PR
XX (CHIR ) CHIRON CORP.
PA
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI: 1989-159274/22.
DR
XX N-PSDB: AAN92097.
DR
XX
XX Purified hepatitis C virus
PT
XX - and associated nucleic acids and polypeptide(s)
XX
XX Claim 13: Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp: English.
XX
XX It is the sequence encoded in the open reading frame of hepatitis C virus
CC cDNA inserts in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,
CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could
CC be used in immunoassay reagents and vaccines and to generate antibodies
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC non-B hepatitis.
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 1766 AA;
SQ
Query Match 99.0%; Score 943; DB 10; Length 1766;
Best Local Similarity 98.4%; Pred. No. 8.8e-89;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCMVYHGA 60
Db 310 LAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCMVYHGA 369
Qy 61 GRTTASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
Db 370 GRTTASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 429
Qy 121 GDSRGSLLSPRISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVSLTETM 180
Db 430 GDSRGSLLSPRISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVSLTETM 489
Qy 181 RS 182
Db 490 RS 491
RESULT 8
AAP90158
XX
XX
XX
DT 25-MAR-2003 (updated)

```

```

ID AAP90158 standard; protein; 1786 AA.
XX
AC AAP90158;
XX
XX 25-MAR-2003 (updated)
DT 10-NOV-1989 (first entry)
XX
XX Protein sequence of hepatitis c virus composite cDNA.
DE
XX
XX Hepatitis C virus; vaccine.
KW
XX
XX Pan troglodytes.
OS
XX
XX GB2212511-A.
PN
XX
XX 26-JUL-1989.
PD
XX
XX 18-NOV-1988; 88GB-0027024.
PF
XX
XX 18-NOV-1987; 87US-0122714.
PR
XX 30-DEC-1987; 87US-0139886.
PR
XX 26-FEB-1988; 88US-0161072.
PR
XX 26-OCT-1988; 88US-0263584.
XX
XX (CHIR ) CHIRON CORPORATION.
PA
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI: 1989-215054/30.
DR
XX N-PSDB: AAN90327.
DR
XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX
XX Disclosure; fig 26; 30pp: English.
XX
XX The sequence is encoded by the composite cDNA of AAN90327. These
CC antigens react with antibodies in patients with non-A non-B hepatitis
CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC antibodies for immunoassay or treatment, or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
XX Sequence 1786 AA;
SQ
Query Match 99.0%; Score 943; DB 10; Length 1786;
Best Local Similarity 98.4%; Pred. No. 8.9e-89;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCMVYHGA 60
Db 310 LAPITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCMVYHGA 369
Qy 61 GRTTASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120
Db 370 GRTTASPKGPVIQMTNVYDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 429
Qy 121 GDSRGSLLSPRISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVSLTETM 180
Db 430 GDSRGSLLSPRISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVSLTETM 489
Qy 181 RS 182
Db 490 RS 491
RESULT 9
AAP90164
ID AAP90164 standard; protein; 2261 AA.
XX
XX AAP90164;
XX
XX 25-MAR-2003 (updated)
DT

```

DT 01-NOV-1989 (first entry)  
 XX  
 DE Peptide encoded by composite hepatitis C virus cDNA.  
 XX  
 KW Hepatitis C virus: clone 12f: clone 15e: probe: vaccine.  
 XX

OS Pan troglodytes.  
 XX

PN GB2212511-A.  
 XX

PD 26-JUL-1989.  
 XX

XX 18-NOV-1988; 88GB-0027024.  
 XX

XX 18-NOV-1987; 87US-0122714.  
 PR

PR 30-DEC-1987; 87US-0139886.  
 PR

PR 26-FEB-1988; 88US-0161072.  
 PR

PR 26-OCT-1988; 88US-0263584.  
 PR

XX (CHIR ) CHIRON CORPORATION.  
 PA

XX Houghton M, Choo QL, Kuo G;  
 XX

XX WPI: 1989-215054/30.  
 XX

DR N-PSDB: AAN90331.  
 DR

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and  
 PT treatment of infection.  
 XX

PS Disclosure: fig 32; 235pp; English.  
 PS

XX The sequence is the peptide encoded by the composite hepatitis C  
 CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose  
 CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,  
 CC or to produce vaccines.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX  
 SQ Sequence 2261 AA;

Query Match 99.0%; Score 943; DB 10; Length 2261;  
 Best Local Similarity 98.4%; Pred. No. 1.2e-88;  
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60  
 :|||||  
 Db 401 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 460  
 :|||||  
 OY 61 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRR 120  
 :|||||  
 Db 461 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRR 520  
 :|||||  
 OY 121 GDSRGSLLSPRPISYKSGSGGPGLLCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 :|||||  
 Db 521 GDSRGSLLSPRPISYKSGSGGPGLLCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 580  
 :|||||  
 OY 181 RS 182  
 :||  
 Db 581 RS 582

RESULT 10  
 AAP92047  
 ID AAP92047 standard; protein: 2301 AA.  
 XX  
 AC AAP92047;

XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)

XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones  
 DE 12f through 15e.  
 XX

KW Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).  
 XX

OS Hepatitis C virus.  
 XX

PN EP318216-A.  
 XX

PD 31-MAY-1989.  
 XX

XX 18-NOV-1988; 88EP-0310922.  
 FF

XX 18-NOV-1987; 87US-0122714.  
 PR

PR 30-DEC-1987; 87US-0139886.  
 PR

PR 26-FEB-1988; 88US-0161072.  
 PR

PR 06-MAY-1988; 88US-0191263.  
 PR

PR 26-OCT-1988; 88US-0263584.  
 PR

PR 14-NOV-1988; 88US-0271450.  
 PR

XX (CHIR ) CHIRON CORP.  
 PA

XX Houghton M, Choo QL, Kuo G;  
 XX

XX WPI: 1989-159274/22.  
 DR

DR N-PSDB: AAN92103.  
 DR

XX Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 PT  
 PS Claim 13; Figure 32-1 - 32-7; 139 pp; English.  
 XX

CC It is the sequence encoded in the open reading frame of hepatitis C virus  
 CC (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could  
 CC be used in immunoassay reagents and vaccines and to generate antibodies  
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,  
 CC non-B hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 CC  
 XX  
 SQ Sequence 2301 AA;

Query Match 99.0%; Score 943; DB 10; Length 2301;  
 Best Local Similarity 98.4%; Pred. No. 1.2e-88;  
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60  
 :|||||  
 Db 401 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 460  
 :|||||  
 OY 61 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRR 120  
 :|||||  
 Db 461 GTRTIASPKGPVIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRR 520  
 :|||||  
 OY 121 GDSRGSLLSPRPISYKSGSGGPGLLCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 :|||||  
 Db 521 GDSRGSLLSPRPISYKSGSGGPGLLCPAGHANGIFRAAVCTRGVAKAVDFIPVESLETTM 580  
 :|||||  
 OY 181 RS 182  
 :||  
 Db 581 RS 582

RESULT 11  
 AAP92050  
 ID AAP92050 standard; protein: 2436 AA.  
 XX  
 AC AAP92050;

XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)

XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones  
 DE K9-1 through 15e.  
 DE  
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)

```

XX OS Hepatitis C virus.
XX PN EP18216-A.
XX XX
XX PD 31-MAY-1989.
XX XX
XX PF 18-NOV-1988; 88EP-0310922.
XX XX
XX PR 18-NOV-1987; 87US-0122714.
XX PR 30-DEC-1987; 87US-0139886.
XX PR 26-FEB-1988; 88US-0161072.
XX PR 06-MAY-1988; 88US-0191263.
XX PR 26-OCT-1988; 88US-0263584.
XX PR 14-NOV-1988; 88US-0271450.
XX XX
XX PA (CHIR ) CHIRON CORP.
XX PI
XX PI Houghton M, Choo QL, Kuo G;
XX DR N-PSDB; AAN92106.
XX DR WPI; 1989-159274/22.
XX XX
XX PT Purified hepatitis C virus
XX PT - and associated nucleic acids and polypeptide(s)
XX PS Claim 13; Figure 47-1 - 47-8; 139 pp; English.
XX CC
XX CC It is the sequence encoded in the open reading frame of hepatitis C virus
XX CC (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
XX CC be used in immunoassay reagents and vaccines and to generate antibodies
XX CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,
XX CC non-B hepatitis.
XX CC (Updated on 25-MAR-2003 to correct PR field.)
XX CC (Updated on 25-MAR-2003 to correct PI field.)
XX SQ Sequence 2436 AA;

Query Match 99.0%; Score 943; DB 10; length 2436;
Best Local Similarity 98.4%; Pred. No. 1.3e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGGLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCTVYHGA 60
DB 576 LAPITAAQOTRGGLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCTVYHGA 635
QY 61 GTRTIASPKGPVIQMTYNDVQDLVGNPAPQGSRSLLPCTCGSSDLVLTTRHADVIPVRRR 120
DB 636 GTRTIASPKGPVIQMTYNDVQDLVGNPAPQGSRSLLPCTCGSSDLVLTTRHADVIPVRRR 695
QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 696 GDSRGSLLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755
QY 181 RS 182
DB 756 RS 757

RESULT 12
AAP90288
ID AAP90288 standard; protein; 2436 AA.
XX AC AAP90288;
XX XX
XX XX 25-MAR-2003 (updated)
DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX XX
XX DE Peptide encoded by composite hepatitis C cDNA.
XX XX
XX KW Hepatitis C virus; clone 15e; clone K9-1; probe; vaccine.
XX XX

```

```

OS Pan troglodytes.
XX XX
XX PN GB2212511-A.
XX XX
XX PD 26-JUL-1989.
XX XX
XX PF 18-NOV-1988; 88GB-0027024.
XX XX
XX PR 18-NOV-1987; 87US-0122714.
XX PR 30-DEC-1987; 87US-0139886.
XX PR 26-FEB-1988; 88US-0161072.
XX PR 26-OCT-1988; 88US-0263584.
XX XX
XX PA (CHIR ) CHIRON CORPORATION.
XX PI
XX PI Houghton M, Choo QL, Kuo G;
XX DR WPI; 1989-215054/30.
XX DR N-PSDB; AAN90336.
XX XX
XX PT Hepatitis C virus gene - used for prodn. of polynucleotide probes,
XX PT polypeptide(s) and antibodies for diagnosis, prevention and
XX PT treatment of infection.
XX PS Disclosure; fig 47-1 to 47-8; 235pp; English.
XX CC
XX CC The sequence is the peptide encoded by the composite hepatitis C
XX CC virus (HCV) cDNA of AAN90336. The polypeptides are used to
XX CC diagnose HCV-induced NANBH, to raise antibodies for
XX CC immunoassay or treatment, or to produce vaccines.
XX CC (N.B. This record was resubmitted to correct errors in the sequence.)
XX CC (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 2436 AA;

Query Match 99.0%; Score 943; DB 10; Length 2436;
Best Local Similarity 98.4%; Pred. No. 1.3e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGGLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCTVYHGA 60
DB 576 LAPITAAQOTRGGLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCTVYHGA 635
QY 61 GTRTIASPKGPVIQMTYNDVQDLVGNPAPQGSRSLLPCTCGSSDLVLTTRHADVIPVRRR 120
DB 636 GTRTIASPKGPVIQMTYNDVQDLVGNPAPQGSRSLLPCTCGSSDLVLTTRHADVIPVRRR 695
QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 696 GDSRGSLLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755
QY 181 RS 182
DB 756 RS 757

RESULT 13
AAB18540
ID AAB18540 standard; Protein; 2772 AA.
XX AC AAB18540;
XX XX
XX XX 15-JAN-2001 (first entry)
DT DT
XX DE Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.
XX XX
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polypeptide;
XX KW viral infectivity; viral replication.
XX XX
XX OS Hepatitis C virus.
XX XX
XX PN EP1034785-A2.
XX XX

```



```
PD 13-SEP-2000.
XX
PF 16-MAR-1990; 2000EP-0109602.
XX
XX 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 18-MAY-1989; 89US-0355002.
PR 16-MAR-1990; 90EP-0302866.
XX
PA (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
PI
XX WPI; 2000-566891/53.
DR N-PSDB; AAN75296.
XX
XX Novel composition comprising a hepatitis C virus antisense
PT polynucleotide which is complementary to or corresponds to a sense
PT strand of the virus genome, and selectively hybridises to it.
XX
XX Example; Fig 16; 75pp; English.
XX
XX The specification describes a pharmaceutical composition which
CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The
CC HCV is characterized by a positive stranded RNA genome which has
CC 40% homology at the polypeptide level to a HCV polypeptide. The
CC antisense polynucleotide binds to cellular polynucleotides which
CC enhance and/or are required for viral infectivity, replicative
CC ability or chronicity. The antisense polynucleotides may also be
CC designed to bind with high specificity, to be of increased stability,
CC to be stable and to have low toxicity. The composition also comprises
CC an agent which causes viral RNA to be inactive. The composition
CC is used for preventing HCV replication in a system. The present
CC sequence is encoded by a novel HCV cDNA sequence, which is used in the
CC course of the invention.
XX
SQ Sequence 2772 AA:
Query Match 99.0%; Score 943; DB 21; Length 2772;
Best Local Similarity 98.4%; Pred. No. 1.6e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
DB 912 LAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 971
QY 61 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSKSLTPCTCGSSDLYLVTRHADVIPVRR 120
DB 972 GTRTIASPKGPVIOMYTNVDQDLVGMWPAPOGSKSLTPCTCGSSDLYLVTRHADVIPVRR 1031
QY 121 GDSRGSLLSPRPISYLGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 1032 GDSRGSLLSPRPISYLGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1091
QY 181 RS 182
DB 1092 RS 1093

RESULT 14
AAR34009
ID AAR34009 standard; Protein; 2816 AA.
XX
XX AAR34009;
AC
XX
XX 25-MAR-2003 (updated)
DT 26-JUL-1993 (first entry)
XX
XX HCV-1 polypeptide.
DE
XX
XX Polymerase chain reaction; PCR; amplifiy; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;

gp53(BVDV)/gp55; hog cholera virus; pestivirus; NSI; flavivirus.
Hepatitis C virus.
WO9306126-A1.
01-APR-1993.
11-SEP-1992; 92WO-US07683.
13-SEP-1991; 91US-0759575.
(CHIR ) CHIRON CORP.
Houghton M, Weiner AJ;
WPI; 1993-117468/14.
Immuno-reactive hepatitis C virus polypeptide compns. - contg.
at least 2 sequences from the first variable domain of distinct
HCV isolates
Disclosure; Fig 9; 106pp; English.
This sequence represents the entire hepatitis C virus polypeptide.
HCV is a member of the flavivirus family and appears to encode a basic
polypeptide domain ("C") at the N-terminal of the viral polypeptide.
followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
nonstructural genes NS2 through NS5. See also AAO39134-48, AAR33982-
4008 and AAR38088-89.
(Updated on 25-MAR-2003 to correct PN field.)
SQ Sequence 2816 AA:
Query Match 99.0%; Score 943; DB 14; Length 2816;
Best Local Similarity 98.4%; Pred. No. 1.6e-88;
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 60
DB 1036 LAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTYHGA 1085
QY 61 GTRTIASPKGPVIOMYTNVDKDLVGMWPAPOGSKSLTPCTCGSSDLYLVTRHADVIPVRR 120
DB 1086 GTRTIASPKGPVIOMYTNVDQDLVGMWPAPOGSKSLTPCTCGSSDLYLVTRHADVIPVRR 1145
QY 121 GDSRGSLLSPRPISYLGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETTM 180
DB 1146 GDSRGSLLSPRPISYLGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205
QY 181 RS 182
DB 1206 RS 1207

RESULT 15
AAR70230
ID AAR70230 standard; protein; 2894 AA.
XX
XX AAR70230;
AC
XX
XX 25-MAR-2003 (updated)
DT 07-NOV-1995 (first entry)
XX
XX Composite hepatitis C virus (HC-J1/CDC/CHI).
XX Composite hepatitis C virus; HC-J1/CDC/CHI; HCV; non-A non-B;
KW synthetic antigens; blood screening.
XX
XX Hepatitis C virus.
OS
XX Ep644202-A1.
PN
XX
```

PD 22-MAR-1995.  
XX  
XX  
PF 14-DEC-1990; 94EP-0108611.  
XX  
XX  
PR 14-DEC-1990; 90EP-0124247.  
PR 14-DEC-1990; 90EP-0108611.  
XX  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
XX  
PI Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;  
XX  
XX WPI: 1995-116946/16.  
DR  
XX  
XX Synthetic antigens for the detection of hepatitis C virus  
PT antibodies - comprise portions of the Hcv peptide sequence, for  
PT use in screening blood and blood products  
XX  
XX  
PS Disclosure; Fig 1; 51pp; English.  
XX  
XX AAR70230 is the composite hepatitis C virus (HC-J1/CDC/CH1) protein  
CC from which the synthetic HCV antigens described in AAR70210-R70229  
CC were derived. These synthetic antigens can be used to screen blood,  
CC or blood products for the presence HCV, they can also be used in  
CC various specific assays for the detection of HCV antibodies, and  
CC antigens, or as immunogens.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
CC (Updated on 25-MAR-2003 to correct PF field.)  
XX  
XX  
SQ Sequence 2894 AA;  
  
Query Match 99.0%; Score 943; DB 16; Length 2894;  
Best Local Similarity 98.4%; Pred. NO. 1.7e-88;  
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MAPITAYAQOTRGLGCIITSITGRKKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 50  
Db :|||||  
1026 LAPITAYAQOTRGLGCIITSITGRKKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 1085  
  
QY 61 GTRTIASPKGPVIOYTNVNDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 120  
Db :|||||  
1086 GTRTIASPKGPVIOYTNVNDKDLVGPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145  
  
QY 121 GDSRGSLLSPRISYILKSSGGPLLCPCAGHAGVIFPRAAVCTRGVAKAVDFIPVESLETTM 180  
Db :|||||  
1146 GDSRGSLLSPRISYILKSSGGPLLCPCAGHAGVIFPRAAVCTRGVAKAVDFIPVENLETTM 1205  
  
QY 181 RS 182  
Db :||  
1206 RS 1207

Search completed: August 30, 2003, 19:12:19  
Job time : 42.2251 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 169.009 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAQTRGLGCIIT.....GVAXAVDFIPVESLETIMRS 182

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlp  
-Q/cqn2.1/USPTO\_spool/US09965594/runat\_29082003.151918.28302/app\_query.fasta\_1.2872  
-DB=N\_Geneseq\_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -ECGN\_1.1\_1412 -runat\_29082003.151918.28302 -NCPU=6 -ICPU=3  
-NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DEEXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSL/cgqdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	953	100.0	549	21	AAA70344 Hepatitis C virus
2	946	99.3	2058	24	AAD29795 HCV-1 NS3/4a mutan
3	946	99.3	2058	24	ABK15344 Hepatitis C virus
4	946	99.3	2058	25	ABX14410 DNA encoding HCV-1
5	943	99.0	5300	10	AAH92097 combined open read
6	943	99.0	5360	10	AAH90327 Hepatitis C virus
7	943	99.0	6905	10	AAH92103 Combined open read
8	943	99.0	7310	10	AAH92106 Composite open read
9	943	99.0	7310	10	AAH90336 Composite hepatitis
10	943	99.0	7310	16	AAH98221 Hepatitis C virus
11	943	99.0	8316	21	AAH75296 cDNA sequence comp
12	943	99.0	9133	20	AAZ07656 Nucleotide sequenc
13	943	99.0	9185	11	AAQ05956 Sense strand of th
14	943	99.0	9185	12	AAQ10566 Hepatitis C virus
15	943	99.0	9185	21	AAH75297 Sense strand of HC
16	943	99.0	9400	13	AAQ21744 Compiled HCV CDNA.
17	943	99.0	9401	17	AAH75297 Hepatitis C virus
18	943	99.0	9401	18	AAH79981 HCV polyprotein co
19	943	99.0	9401	19	AAV09989 HCV polyprotein co
20	943	99.0	9401	24	AAD35043 Hepatitis C virus
21	942	98.8	2061	24	AAD34500 Hepatitis C virus
22	942	98.8	2061	24	AAD31767 Hepatitis C virus
23	940	98.6	9185	20	AAZ26737 Nucleotide sequenc
24	940	98.6	9185	20	AAH00459 Hepatitis C virus
25	939	98.5	8316	11	AAQ05955 Hepatitis C virus
26	939	98.5	9502	15	AAQ74770 Hepatitis C virus
27	937	98.3	9646	19	AAV59361 Hepatitis C virus
28	937	98.3	9646	24	ABK87285 CDNA encoding hepa
29	937	98.3	12980	19	AAV59364 Hepatitis C virus
30	937	98.3	12980	24	ABK87286 Hepatitis C virus
31	937	98.3	16622	21	AAZ36212 Nucleotide sequenc
32	936	98.2	630	17	AAT43708 Plasmid pT5H1s/HIV
33	936	98.2	630	17	AAT42394 HCV insoluble NS3
34	936	98.2	630	18	AAT58401 HCV NS3 protease C
35	936	98.2	810	17	AAT38903 PNB182delta4A HT e
36	936	98.2	810	17	AAT42389 HCV solubilised NS
37	936	98.2	810	18	AAT58396 HCV soluble NS3 pr
38	935	98.1	6299	22	AAF83669 HCV NS3A ORF comp
39	934	98.0	1933	20	AAZ33258 HCV NS3 DNA. Hepa
40	934	98.0	8145	20	AAZ32529 Plasmid pET-BS(+)/
41	933	97.9	594	21	AAZ73335 Hepatitis C virus
42	933	97.9	9365	24	AAD25518 Hepatitis C virus
43	933	97.9	9401	17	AAH41882 Hepatitis C virus
44	933	97.9	9416	19	AAV59378 Hepatitis C virus
45	933	97.9	9416	24	ABK87300 CDNA encoding hepa

ALIGNMENTS

RESULT 1  
AAA70344  
ID AAA70344 standard; DNA; 549 Bp.  
XX  
XX AAA70344;  
AC  
XX  
DT 19-DEC-2000 (first entry)  
XX  
XX Hepatitis C virus NS3 protease coding sequence.  
DE  
DE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; ds.  
XX  
OS Hepatitis C virus.  
XX  
XX Key Location/Qualifiers  
FT CDS 1..549



## Alignment Scores:

```

Pred. No.:      1,46e-78      Length:      2058
Score:          946.00      Matches:      180
Percent Similarity: 100.00%      Conservatives: 2
Best Local Similarity: 98.90%      Mismatches: 0
Query Match:      99.27%      Indels: 0
DB:              24      Gaps: 0

US-09-965-594-1 (1-182) x AAD29795 (1-2058)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1 ATGGCGCCCATCACGGCGTACGCCAGCAGACAAGGGCCCTCTAGGGTGCAATATCACC 60
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGGCGGGGACAAACCAACCACTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGlyVala 60
Db 121 GCCCAAACTTCTCGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAACGAGGACCATCGCGTCACCCAAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGCTGGCGCGCTCCCAAGGTACCCGATCATTGACACCCCTGCACCTGC 300
QY 101 GlySerSerAspLeuThrValThrArgHisAlaAspValIleProValIleArgArgArg 120
Db 301 GGCTCTCGGACCTTACCTGGTCTACAGAGCAGCGCGGATGATTCCTCTGAAAGGCTCTCTG 420
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProLysCysTrpThrValIleHisGly 140
Db 361 GGTATAGCAGGGGAGCGCTGTGTGCGCCGCGCCATTTCTACTTGAAGGCTCTCTG 480
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 421 GGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTTAGGCGCGCGGTGTGTC 480
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACAACCAATG 540

QY 181 ArgSer 182
Db 541 AGGTCC 546

RESULT 3
ABK15344
ID ABK15344 standard; DNA; 2058 BP.
XX AC ABK15344;
XX AC
XX AC
DT 08-MAY-2002 (first entry)
XX
XX Hepatitis C virus NS3/4a conformational epitope gene sequence.
XX
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW Immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; gene; ds.
XX
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH 1..2058
FT CDS /*tag- a
FT /partial
FT /product- "HCV NS3/4a conformational epitope"
FT /note- "This sequence lacks a stop codon"
XX

```

```

PN W0200196870-A2.
XX
PD 20-DEC-2001.
XX
PF 14-JUN-2001; 2001WO-US19156.
XX
PR 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
XX
DR WPI; 2002-090228/12.
DR P-PSDB; AAU76377.
XX
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support .
XX
PS Disclosure; Fig 3; 92pp; English.
XX
XX The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present nucleic acid sequence
CC encodes the non-structural protein NS3/4a conformational epitope of the
CC invention.
XX
SQ Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

Alignment Scores:
Pred. No.:      1,46e-78      Length:      2058
Score:          946.00      Matches:      180
Percent Similarity: 100.00%      Conservatives: 2
Best Local Similarity: 98.90%      Mismatches: 0
Query Match:      99.27%      Indels: 0
DB:              24      Gaps: 0

US-09-965-594-1 (1-182) x ABK15344 (1-2058)

QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1 ATGGCGCCCATCACGGCGTACGCCAGCAGACAAGGGCCCTCTAGGGTGCAATATCACC 60
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 61 AGCCTAACTGGCGGGGACAAACCAACCACTGGAGGTGAGGTCCAGATTGTGTCAACTGCT 120
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGlyVala 60
Db 121 GCCCAAACTTCTCGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACACGGGGCC 180
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 181 GGAACGAGGACCATCGCGTCACCCAAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 240
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 241 CAAGACCTTGTGGCTGGCGCGCTCCCAAGGTACCCGATCATTGACACCCCTGCACCTGC 300
QY 101 GlySerSerAspLeuThrValThrArgHisAlaAspValIleProValIleArgArgArg 120
Db 301 GGCTCTCGGACCTTACCTGGTCTACAGAGCAGCGCGGATGATTCCTCTGAAAGGCTCTCTG 420
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProLysCysTrpThrValIleHisGly 140
Db 361 GGTATAGCAGGGGAGCGCTGTGTGCGCCGCGCCATTTCTACTTGAAGGCTCTCTG 480
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 421 GGGGGTCCGCTGTGTGCCCCCGGGGACGCGGTGGGCATATTTAGGCGCGCGGTGTGTC 480
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 481 ACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACAACCAATG 540

QY 181 ArgSer 182
Db 541 AGGTCC 546

RESULT 3
ABK15344
ID ABK15344 standard; DNA; 2058 BP.
XX AC ABK15344;
XX AC
XX AC
DT 08-MAY-2002 (first entry)
XX
XX Hepatitis C virus NS3/4a conformational epitope gene sequence.
XX
XX Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW Immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; gene; ds.
XX
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH 1..2058
FT CDS /*tag- a
FT /partial
FT /product- "HCV NS3/4a conformational epitope"
FT /note- "This sequence lacks a stop codon"
XX

```

QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
 Db :::  
 241 CAAGACCTTGCGGCTGGCCGCTCCGCAAGGTAGCCGATCATTCACACCTGCCTTGC 300  
 QY 101 GlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 120  
 Db :::  
 301 GGCTCTCGGACCTTTACCTGGTCACGAGGCGCGCATGTCTCCCTGCGCGCGG 360  
 QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140  
 Db :::  
 361 GGTGATAGCAGGCGGACCTGCTGCGCGCGCGCATTTCTACTTGAAGGCTCCTCG 420  
 QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValCys 160  
 Db :::  
 421 GGGGCTCCGCTGTGTGCCCCGGGCGCGCGCATTTAGGCGCGGCTGTGC 480  
 QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMet 180  
 Db :::  
 481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACCACTG 540  
 QY 181 ArgSer 182  
 Db ::::::::::  
 541 AGGTCC 546

RESULT 4  
 ABX14410  
 ID ABX14410 standard; DNA; 2058 BP.  
 XX  
 AC ABX14410;  
 XX  
 DT 06-MAR-2003 (first entry)  
 XX  
 DE DNA encoding HCV-1 NS3/4a conformational antigen.  
 XX  
 KW Immunoassay solid support; Hepatitis C virus type-1; HCV-1;  
 KW NS3/4a conformational epitope; multiple epitope fusion antigen;  
 KW HEFA; anti-HCV antibody; NS3/4a conformational antigen;  
 KW HCV infection; mutant; gene; ds.  
 XX  
 OS Hepatitis C virus type 1.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT CDS 1..2058  
 FT /tag- a  
 FT /partial  
 FT /product- "NS3/4a conformational antigen"  
 FT /note- "This sequence lacks a stop codon"  
 XX  
 PN US2002146685-A1.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 14-JUN-2001; 2001US-0881654.  
 XX  
 PR 15-JUN-2000; 2000US-212082P.  
 PR 02-APR-2001; 2001US-280811P.  
 PR 02-APR-2001; 2001US-280867P.  
 XX  
 XX (CHIE/) CHEN D Y.  
 PA (ARCA/) ARCANDEL P.  
 PA (TAND/) TANDESKE L.  
 PA (GEOR/) GEORGE-NASCIMENTO C.  
 PA (COIT/) COIT D.  
 PA (MEDI/) MEDINA-SELBY A.  
 XX  
 PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
 PI Medina-Selby A;  
 XX  
 XX WPI: 2003-147573/14.  
 DR P-PSDB; ABG72261.  
 DR  
 XX

PT Immunoassay solid support for detecting Hepatitis C Virus infection in  
 PT biological samples, comprises Hepatitis C Virus conformational epitope  
 PT and multiple epitope fusion antigen -  
 XX  
 PS Disclosure; Fig 3A-3D; 45pp; English.  
 XX  
 CC The present invention relates to immunoassays comprising Hepatitis C  
 CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
 CC antigen (HEFA), bound to a solid support. The NS3/4a epitope and/or  
 CC the multiple epitope fusion antigen react with anti-HCV antibodies  
 CC present in a biological sample from an HCV-infected individual. The  
 CC immunoassays and methods of the invention are useful for detecting  
 CC HCV infection in a biological sample. The inventive immunoassay solid  
 CC support provides a sensitive and reliable method for detecting early  
 CC HCV seroconversion. The assays can detect HCV infection caused by any  
 CC six known genotypes of HCV. The use of the multiple epitope fusion  
 CC proteins decreases masking problems, improves sensitivity in detecting  
 CC antibodies by allowing a greater number of epitopes on a unit area  
 CC of substrate, and improves selectivity. The present sequence  
 CC encodes HCV type 1 (HCV-1) NS3/4a conformational antigen, a mutant of  
 CC the HCV-1 NS3/4a polypeptide.

SQ Sequence 2058 BP; 419 A; 633 C; 581 G; 425 T; 0 other;

## Alignment Scores:

Pred. No.: 1,46e-78 Length: 2058  
 Score: 946.00 Matches: 180  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 98.90% Mismatches: 0  
 Query Match: 99.27% Indels: 0  
 DB: 25 Gaps: 0

US-09-965-594-1 (1-182) x ABX14410 (1-2058)

QY 1 MetalProIleThrAlaTyrAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
 Db 1 ATGGCGCCCATCAGCGGTACGCCAGCAGACAGAGGGGCTCTCTAGGTGCTAATCACC 60  
 QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThrAla 40  
 Db 61 AGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACTGCT 120  
 QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGlyAla 60  
 Db 121 GCCCAACCTTCTCGGCAACGTGCATCAATGGGCTGTCTGGACTGTCTACACGGGGC 180  
 QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80  
 Db 181 GGAACGAGGACCATCGCTCACCAAGGTCCTGTCTATCCAGATGTATACCAATGTAGAC 240  
 QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
 Db 241 CAAGACCTTGCGGCTGGCCGCTCCGCAAGGTAGCCGATCATTCACACCTGCCTTGC 300  
 QY 101 GlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 120  
 Db 301 GGCTCTCGGACCTTTACCTGGTCACGAGGCGCGCATGTCTCCCTGCGCGCGG 360  
 QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 140  
 Db 361 GGTGATAGCAGGCGGACCTGCTGCGCGCGCGCATTTCTACTTGAAGGCTCCTCG 420  
 QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaValCys 160  
 Db 421 GGGGCTCCGCTGTGTGCCCCGGGCGCGCGCATTTAGGCGCGGCTGTGC 480  
 QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrMet 180  
 Db 481 ACCCGTGGAGTGGCTAAGGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACCACTG 540  
 QY 181 ArgSer 182  
 Db 541 AGGTCC 546

## RESULT 5

AAN92097  
 ID AAN92097 standard; DNA; 5300 BP.  
 AC  
 AC AAN92097;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX  
 XX Combined open reading frames of the hepatitis C virus (HCV) cDNA in  
 DE clones 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c,  
 DE 8f, 33f, 33g and 39c.  
 DE  
 DE Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.  
 XX  
 XX Hepatitis C virus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 3..5300  
 FT CDS /\*tag= a  
 FT  
 FT  
 PN EP318216-A.  
 XX  
 XX 31-MAY-1989.  
 PD  
 XX 18-NOV-1988; 88EP-0310922.  
 PF  
 XX 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Houghton M, Choo QL, Kuo G;  
 PI  
 XX WPI: 1989-159274/22.  
 DR P-PSDB; AAP92041.  
 DR  
 XX Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 PT  
 XX Claim 3; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.  
 PS  
 XX It is a double-stranded nucleotide sequence of the open reading frame  
 CC (ORF) (tag a) extending through clones 14i, 11b, 7f, 7e, 8h, 33c, 40b,  
 CC 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c of hepatitis C  
 CC virus (HCV) cDNA. In creating the composite sequence the following  
 CC heterogeneities were considered. Clone 33c contains a sequence  
 CC of 800 base pairs which overlaps the cDNAs in clones 40b and 37c. In  
 CC clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is  
 CC a G. However, in clone 37b the corresponding nucleotide is an A. This  
 CC heterogeneity may have important ramifications for protein folding.  
 CC Nucleotide #2 in clone 8h is a T which may represent a cloning artifact  
 CC because the corresponding residue in clone 7e and in 3 other overlapping  
 CC clones is an A. Therefore the residue in this position is designated as  
 CC an A. The 3'-terminal nucleotide in clone 8f is represented as a T  
 CC than a G because the corresponding residue in clone 33f and in 2 other  
 CC overlapping clones is a T. The 3' terminal sequence of clone 33f is  
 CC represented as ATTC, as is found in the corresponding sequence in clone  
 CC 33g and in 2 other overlapping clones, rather than as TTGC, as is found  
 CC in clone 33f. Residue #4 in clone 33g is designated as A rather than a T  
 CC because the corresponding residue in clone 33f and 2 other overlapping  
 CC clones is an A. The 3'-terminus of clone 14i is depicted as TA rather  
 CC than AA because the corresponding dinucleotide in clone 11b and 3 other  
 CC clones is TA. Potential cloning artifacts have been omitted and instead  
 CC the corresponding sequences in non-5'-terminal regions of multiple  
 CC overlapping clones are shown. AAN92097 could be used as a source of  
 CC oligomeric DNA hybridisation probes to detect the presence of HCV  
 CC nucleic acids in samples. The polypeptide(s) it encodes could be used as

CC immuno- assay reagents and vaccines and to generate antibodies useful in  
 CC diagnosis and passive immunotherapy for HCV infection/non-A, non-B  
 CC hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 5300 BP; 1047 A; 1606 C; 1515 G; 1130 T; 2 other;  
  
 Alignment Scores: 8.9e-78 Length: 5300  
 Pred. No.: 943.00 Matches: 179  
 Score: 100.00% Conservative: 3  
 Percent Similarity: 98.35% Mismatches: 0  
 Best Local Similarity: 98.95% Indels: 0  
 Query Match: 10 Gaps: 0  
 DB: 10  
  
 US-09-965-594-1 (1-182) x AAN92097 (1-5300)  
  
 QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleThr 20  
 DB 930 CTGGCGCCCATCAGCGGTAGCCCGCAGCAGACAGGGGCGCTCTAGGTGCATAATCACC 989  
  
 QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40  
 DB 990 AGCCTAACTGGCGGGACAAAACCAAGTGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049  
  
 QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValThrHisGlyAla 60  
 DB 1050 GCCCAACCTTCTGGCAACGTGCATCAATGGGTGTGCTGGACTGTCTACACGGGGCC 1109  
  
 QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValasp 80  
 DB 1110 GGAACGAGGACCATCGGTCAACCAAGGGTCTCTGCATCCAGATGTATACCAATGTAGAC 1169  
  
 QY 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
 DB 1170 CAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCACTTGC 1239  
  
 QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120  
 DB 1230 GGCTCTCGGACCTTTACCTGGTGCAGGAGCACGCCGATGTCTATCCCGTGGCGGGGG 1289  
  
 QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140  
 DB 1290 GGTGATAGAGGGGCGACCTGCTGCGCCCGCCGCGCCATTTCTACTTGAAGGCTCCCTCG 1349  
  
 QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValAlaGlyIlePheArgAlaAlaValCys 160  
 DB 1350 GGGGGTCCGCTGTGTGCCCCCGGGGCGACGCCGTGGGCATATTTAGGCGCGGTGTGC 1409  
  
 QY 161 ThrArgGlyValAlaLysAlaValaspPheIleProValGluSerLeuGluThrThrMet 180  
 DB 1410 ACCCGTGGAGTGCGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGAACCACTG 1469  
  
 QY 181 ArgSer 182  
 DB 1470 AGGTCC 1475  
  
 RESULT 6  
 AAN90327  
 ID AAN90327 standard; cDNA; 5360 BP.  
 XX  
 AC AAN90327;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 11-NOV-1989 (first entry)  
 XX  
 XX Hepatitis C virus composite probe.  
 DE  
 XX Hepatitis C virus; composite cDNA; probe; vaccine.  
 KW  
 XX Pan troglodytes.  
 OS  
 XX

FH	Key	Location/Qualifiers
CDS		3..5360
FT		/tag- a
XX		
PX		
PN		
XX	GB212511-A.	
XX	26-JUL-1989.	
XX		
XX	18-NOV-1988:	88GB-0027024.
XX		
XX	18-NOV-1987:	87US-0122714.
PR	30-DEC-1987:	87US-0139886.
PR	26-FEB-1988:	88US-0161072.
PR	26-OCT-1988:	88US-0263584.
XX		
PA	(CHIR ) CHIRON CORPORATION.	
XX		
PI	Houghton M, Choo QL, Kuo G;	
XX		
WPI	1989-215054/30.	
XX		
PT	Hepatitis C virus gene - used for prodn. of polynucleotide probes,	
PT	polypeptides(s) and antibodies for diagnosis, prevention and treatment	
PT	of infection.	
XX		
PS	Disclosure; Fig. 26; 174pp; English.	
XX		
CC	The sequence shows the composite cDNA sequence derived from the aligned	
CC	hepatitis C virus (HCV) cDNA's in clones 14i, 11b, 7f, 7e, 8h, 3c, 40b,	
CC	37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA, 40b,	
CC	encodes antigens which react with antibodies in patients with non-A	
CC	non-B hepatitis (NANBH). The cDNA can be used to design probes, or to	
CC	synthesize polypeptides, which are used to diagnose HCV-induced, or to	
CC	to raise antibodies for immunoassay or treatment, or to produce	
CC	vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.	
CC	(Updated on 25-MAR-2003 to correct PR field.)	
XX		
SQ	Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;	
 Alignment Scores:		
Pred. No.:	9.03e-78	Length: 5360
Score:	943.00	Matches: 179
Percent Similarity:	100.00%	Conservative: 3
Best Local Similarity:	98.35%	Mismatches: 0
Query Match:	98.95%	Indels: 0
DB:	10	Gaps: 0
 US-09-965-594-1 (1-182) x AAN90327 (1-5360)		
QY	1 MetAlaProIleThrAlaTyrrAlaGlnInThrArgGlyLeuLeuCysIleIleThr 20	
Dd	::::CTGGCGCCATCAGCGGTACGCCAGCACAAAGGGGCTCTTAGGGTGCAATAACACC 989	
QY	21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThraLa 40	
Dd	990 AGCCTAACTTGC CGCGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGTCAACTGCT 1049	
QY	41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrrHisGlyAla 60	
Dd	1050 GCCCAAACCTTCCTGGCAACGTCATCAATGGGCTGTGCTGGACTGTCTACCAACGGGGCC 1109	
QY	61 GlyThrArgThrIleAlaSerProLysGlyProValIIGlnMetTyrrThrAsnValasp 80	
Dd	1110 GGAAACGAGGACCATCGGTGTCACCAAGGUTCCGTGTCATCCAGAATGATACCAATGTAGAC 1169	
QY	81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuthrProCysThrCys 100	
Dd	1170 CAAGACCTTCTGGCTGGCCCGCTCCGCCAAGGTACCGGCTCATTGACACCC:GCACCTGC 1229	
QY	101 GlysSerSerAspLeuTyrrLeuValThrArgHisAlaAspValIleProValArgArgArq 120	
Dd	1230 GGGTCCTCGGACCTTTACCTGGTCACGAGGACGCCGATGTCATTCCCGTGC CGCGGCGG 1289	



	XX	PF	18-NOV-1988;	88EP-0310922.	
	XX	PR	18-NOV-1987;	87US-0122714.	
	PR	30-DEC-1987;	87DS-0139886.		
	PR	26-FEB-1988;	88US-0161072.		
	PR	06-MAY-1988;	88US-0191263.		
	PR	28-OCT-1988;	88US-0263584.		
	PR	14-NOV-1988;	88US-0271450.		
	XX	PA	(CHIR ) CHIRON CORP.		
	XX	PI	Houghton M, Choo QL, Kuo G;		
	XX	PI	WPI; 1989-159274/22.		
	DR	P-PSDB; AAP92050.			
	XX	PT	Purified hepatitis C virus		
	PT	- and associated nucleic acids and polypeptide(s)			
	XX	PS	Claim 3; Figure 47-1 - 47-8; 139pp; English.		
	XX	CC	It is a double-stranded nucleotide sequence of the open reading frame (ORF) (tag a) extending through clones K9-1 to 15e of hepatitis C virus (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to detect the presence of HCV nucleic acids in samples. The polypeptide(s) it encodes could be used as immunocassay reagents and vaccines and to generate antibodies useful in diagnosis and passive immunotherapy for HCV infection/non-A, non-B hepatitis.		
	CC	CC	(Updated on 25-MAR-2003 to correct PI field.)		
	CC	CC	(Updated on 25-MAR-2003 to correct PI field.)		
	XX	SQ	Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other:		
		Alignment Scores:			
		Pred. No.:	1,32E-77	Length:	7310
		Score:	943.00	Matches:	179
		Percent Similarity:	100.00%	Conservative:	3
		Best Local Similarity:	98.35%	Mismatches:	0
		Query Match:	98.95%	Indels:	0
		DB:	10	Gaps:	0
		US-09-965-594-1 (1-182) x AAN92106 (1-7310)			
	Qy	1	MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuCysIleIleThr	20	
	Dd	1728	CTGGCGCCATCATCGGCGTACGCCAGCAGCAAGGGGCCTCCTAGGTGCATAATACC	1787	
	Qy	21	SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThraLa	40	
	Dd	1788	AGCCTAACTGCGCGGACAAAACCAAGTGAGGTGGGTGCTGCTGGACTGTCTAACACTGC	1847	
	Qy	41	AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGlyAla	60	
	Dd	1848	GCCCAAACCTTCCTGGCAACGTGCATCAATGGGTGGTGTCTGGACTGTCTACACGGGCC	1907	
	Qy	61	GlyThrArgThrIleAlaserProLysGlyProValIleGlnMetTyThrAsnValasp	80	
	Dd	1908	GGAACGAGGACCATCGCTCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC	1967	
	Qy	81	LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys	100	
	Dd	1968	CAGACCTTTGGGCTGGCCGCTCGCAAGTAGCCGCTCATTCACACCTGCACCTTGC	2027	
	Qy	101	GlySerSerAspLeuTyrrLeuValThrArgHisAlaAspValIleProvalArgArg	120	
	Dd	2028	GGCTCCTCGACCTTTACCTGGTGCAGAGGCACGCCGATGTCTATTCCGCTGCGCGGG	2087	
	Qy	121	GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrrLeuLysGlySerSer	140	
	Dd	2088	GGTGATAGAGGGGACGCTGTGTGCCCCGCCCATTTCTACTTTGAAGGCTCTCTCG	2147	
	Qy	141	GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys	160	
		RESULT 8			
	ID	AAN92106			
	ID	AAN92106 standard; DNA: 7310 BP.			
	AC	AAN92106;			
	XX	25-MAR-2003 (updated)			
	DT	02-MAR-1990 (first entry)			
	DE	Combined open reading frames of the hepatitis C virus (HCV) cDNAs from clones K9-1 through 15e.			
	DE	Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.			
	OS	Hepatitis C virus.			
	FH	Key Location/Qualifiers			
	FT	CDS 3..7310			
	FT	/*tag= a			
	PN	EP18216-A.			
	PD	31-MAY-1989.			

```

|||||
Db 2148 GGGGTCCTGTGTGCCCCGGGACAGCGGTGGGCAIATTTAGGGCCGGGTGTC 2207
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCGGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAAGCTAGAGACACCATG 2267
Qy 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 9
ID AAN90336 standard; DNA: 7310 BP.
AC AAN90336;
XX 25-MAR-2003 (updated)
DT 19-JUL-2001 (updated)
DT 01-NOV-1989 (first entry)
XX Composite hepatitis C virus (HCV) cDNA.
XX Hepatitis C virus; cDNA; clone 15e; clone k9-1; probe; vaccine; ds.
XX Pan troglodytes.
XX GB2212511-A.
XX 26-JUL-1989.
XX 18-NOV-1988; 88GB-0027024.
XX 18-NOV-1987; 87US-0122714.
XX 30-DEC-1987; 87US-0139886.
XX 26-FEB-1988; 88US-0161072.
XX 26-OCT-1988; 88US-0263584.
XX (CHIR) CHIRON CORPORATION.
XX Houghton M, Choo QL, Kuo G;
XX WPI: 1989-215054/30.
XX P-PSDB; AAP90288.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
XX polypeptide(s) and antibodies for diagnosis, prevention and treatment
XX of infection.
XX Disclosure; fig 47; 235pp; English.
XX The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by
XX aligning clones k9-1 through 15e in 5'-3' direction. The cDNA
XX encodes antigens which react with antibodies in patients with non-A
XX non-B hepatitis (NANBH). The cDNA can be used to design probes, or to
XX synthesise polypeptides, which are used to diagnose HCV-induced
XX to raise antibodies for immunoassay or treatment, or to produce
XX vaccines. See also AAP90288, and AAN90303-35.
XX (N.B. This record was resubmitted to correct errors in the sequence.)
XX CC (Updated on 25-MAR-2003 to correct PR field.)
SQ Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;

Alignment Scores:
Pred. No.: 1,32e-77 Length: 7310
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-1 (1-182) x AAN90336 (1-7310)

```

```

Cy 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 1728 CTGGCCCATCATCAGGGTACGCCACAGCAGACAGAGGGGCTCTTAGGGTGCATATACCC 1787
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 1788 AGCTAACTGGCGGGACAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACTGCT 1847
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 1848 GCCCAACCTTCTTGGCAACGTGCATCAATGGGGTGTGTGGACTGCTTACCACGGGGCC 1907
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 1908 GGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTCATCCAGATGTATACCAATGTAGAC 1967
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 1968 CAAGACCTTGTGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCATTGC 2027
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg 120
Db 2028 GGCTCTCGGACCTTTACTGTGTCCAGGACGACGCCGATGTCAATCCCGCGCGGCGG 2087
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 2088 GGTGATAGCAGGGGACGCTGCTGCGCCCGGCCCATTTCTCTACTTGAAGGCTCCTCG 2147
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 2148 GGGGTGCGCTGTGTGCCCCGGGGCACGCCGTGGGCATATTTAGGCCCGCGGTGTC 2207
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 2208 ACCGGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAAGCTAGAGACACCATG 2267
Qy 181 ArgSer 182
Db 2268 AGGTCC 2273

RESULT 10
ID AAN98221 standard; cDNA to mRNA; 7310 BP.
XX AAQ98221;
AC AAQ98221;
XX 25-MAR-2003 (updated)
DT 15-AUG-1996 (first entry)
XX Hepatitis C virus clone genome.
XX Hepatitis C virus; HCV; antigen; detection; diagnosis; vaccine;
XX antibodies; immunoprophylaxis; sera; serum; ds.
XX Hepatitis C virus.
XX US5443965-A.
XX 22-AUG-1995.
XX 05-APR-1991; 91US-0681703.
XX 06-APR-1990; 90US-0505611.
XX 09-OCT-1990; 90US-0594854.
XX (GENE-) GENELABS INC.
XX Kim JP, Moockli R, Reyes GR;
XX WPI: 1995-302120/39.
XX New nucleic acids encoding hepatitis C virus antigens - used to
XX develop prods. for detection of HCV-infected sera and prodn. of
PT

```

PT vaccines and anti-HCV antibodies.

PS Example 4; Figure 11; 71pp; English.

XX Hepatitis C virus (HCV) antigens can be used for detecting HCV  
 CC infected sera and individuals infected with HCV. They can also be  
 CC used in an anti-HCV vaccine or for the production of anti-HCV  
 CC antibodies which can be used for passive immunoprophylaxis. The  
 CC antigens consistently identify more HCV positive serum samples with  
 CC a high degree of specificity. See AAQ98202-14 and AAR81939-51.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 CC (Updated on 25-MAR-2003 to correct pr field.)

XX Sequence 7310 BP; 1494 A; 2217 C; 2060 G; 1539 T; 0 other;  
 SQ

#### Alignment Scores:

Pred. No.: 1.32e-77 Length: 7310  
 Score: 943.00 Matches: 179  
 Percent Similarity: 100.00% Conservativity: 3  
 Best Local Similarity: 98.35% Mismatches: 0  
 Query Match: 98.95% Indels: 0  
 DB: 16 Gaps: 0

US-09-965-594-1 (1-182) x AAQ98221 (1-7310)

Qy 1 MetAlap-roilethrAlaTyAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20  
 Db 1728 CTGGCGCCCATCAGCGGTAGCCAGCAGCAGAGGGGCTCTTAGGGTGCATATACACC 1787  
 Qy 21 SerLeuThrGlyArgAspIleAsnGlnValGluGlyGluValGlnIleValSerIleAla 40  
 Db 1788 AGCCTTAACGTGGCGGCACAAACCAAGTGCAGGTGAGTCCAGATGTGTCAACPGCT 1847  
 Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlnValCysIleThrValIleValGlyAla 60  
 Db 1848 GCCCAACACTTCTCGCAACGTGCATCAATGGGTGCTGAGCTGTCTACCCAGGGGCC 1907  
 Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnValAsp 80  
 Db 1908 GGAACGAGGACCATCGCGTCCACCAAGGTCTCTCATCCAGATGTATACCAATGTAGAC 1967  
 Qy 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100  
 Db 1969 CAAGACCTTGTGGGTGGCGGTCCGCAAGGTAGCGGCTCATTCACACCTGCACTGC 2027  
 Qy 101 GlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 120  
 Db 2028 GGTCTCTCGGACCTTACTGTGTCAGAGGACCGCGATGTCATTCCGTGCGCGCGG 2087  
 Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerIleThrValGlySerSer 140  
 Db 2088 GGTGTAGCAGGGGAGCGCTGCTGCGCGCGGCGCCATTCTCTACTTGAAGGCTCTCG 2147  
 Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160  
 Db 2148 GGGGTGCTCGCTGTGTGCGCGCGGGGACCGCGTGGGCATATTAGCGCGCGGTGTC 2207  
 Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180  
 Db 2208 ACCGTGGAGTGGCTAAGCGGTGACCTTATCTCCCTGTGGAGAACCTTAGACACCACTG 2267  
 Qy 181 ArgSer 182  
 Db 2268 AGGTCC 2273  
 RESULT 11  
 ID AAA75296 standard; cDNA; 8316 BP.  
 AC AAA75296:  
 XX  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX

DE cDNA sequence compiled Hepatitis C virus cDNA clones.

XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;  
 KW viral infectivity; viral replication; ds.

OS Hepatitis C virus.

PH Key Location/Qualifiers  
 FT CDS 1..8316  
 FT /tag= a  
 FT /note= "partial sequence; no termination codon given"

XX EP1034785-A2.

XX 13-SEP-2000.

XX 16-MAR-1990; 2000EP-0109602.

XX 17-MAR-1989; 89US-0325338.

XX 20-APR-1989; 89US-0341334.

XX 18-MAY-1989; 89US-0355002.

XX 16-MAR-1990; 90EP-0302866.

XX (CHIR ) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI: 2000-566891/53.

XX P-PSDB; AAB18540.

XX Novel composition comprising a hepatitis C virus antisense  
 PT polynucleotide which is complementary to or corresponds to a sense  
 PT strand of the virus genome, and selectively hybridizes to it -

XX Example; Fig 16; 75pp; English.

XX The specification describes a pharmaceutical composition which  
 CC comprises a hepatitis C virus (HCV) antisense polynucleotide. The  
 CC HCV is characterized by a positive stranded RNA genome which has  
 CC 40% homology at the polypeptide level to a HCV polyprotein. The  
 CC antisense polynucleotide binds to cellular polynucleotides which  
 CC enhance and/or are required for viral infectivity, replicative  
 CC ability or chronicity. The antisense polynucleotides may also be  
 CC designed to bind with high specificity, to be of increased stability,  
 CC to be stable and to have low toxicity. The composition also comprises  
 CC an agent which causes viral RNA to be inactive. The composition  
 CC is used for preventing HCV replication in a system. The present  
 CC sequence represents a novel HCV cDNA sequence, which is used in the  
 CC course of the invention.

XX SQ Sequence 8316 BP; 1671 A; 2529 C; 2345 G; 1771 T; 0 other;

Alignment Scores:

Pred. No.: 1.55e-77 Length: 8316  
 Score: 943.00 Matches: 179  
 Percent Similarity: 100.00% Conservativity: 3  
 Best Local Similarity: 98.35% Mismatches: 0  
 Query Match: 98.95% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-1 (1-182) x AAA75296 (1-8316)

Qy 1 MetAlap-roilethrAlaTyAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20

Db 2734 CTGGCGCCCATCAGCGGTAGCCAGCAGCAGAGGGGCTCTTAGGGTGCATATACACC 2793

Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerIleAla 40

Db 2794 AGCCTTAACGTGGCGGCACAAACCAAGTGGAGGTGAGTCCAGATGTGTCAACCTGCT 2853

Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValThrHisGlyAla 60

Db 2854 GCCCAACACTTCTCGCAACGTGCATCAATGGGTGCTGCTGACCTGTCTACCAACGGGCC 2913

```

Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 2914 GGAACGAGGACCATCGCTCCACCAAGGCTCTGTCATCCAGATGTATACCAATGTAGAC 2973
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 2974 CAAGACCTGTGGCTGGCCGCTCGGCAAGGTAGCCCTCATTCACACCCCTGCCTTGC 3033
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
Db 3034 GGCTCTCGACCTTACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3093
Qy 121 GlyAspSerArgGlySerLeuLeuSerProAcqProIleSerTyrLeuLysGlySer 140
Db 3094 GGTGATAGGAGGCGGACCTGCTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGG 3153
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 3154 GGGGCTCGCTGTGTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3213
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGlnSerLeuCluThrThrMet 180
Db 3214 ACCCGTGGAGTGGCTAAGCGGGTGGACTTTATCCCTGTGGAGACCTAGAGACACCATG 3273
Qy 181 ArgSer 182
Db 3274 AGGTCC 3279
RESULT 12
AAZ07656
ID AAZ07656 standard; DNA: 9133 BP.
XX
AC AAZ07656:
XX
XX 20-MAR-2003 (updated)
DT 08-NOV-1999 (first entry)
XX
DE Nucleotide sequence of HCV-1 ORF.
XX
XX Hepatitis C virus; HCV; J1; J7; HCV-1: non-A, non-B HCV; NANBH;
KW HCV infection; vaccine; ds.
XX
OS Hepatitis C virus.
XX
FH Key location/Qualifiers
FT CDS 268..9132
FT /*Tag= b
FT /trans= except- (pos:1588..1589; aa:Leu)
FT /note= "this codon has an apparent 1 nucleotide deletion,
FT which alters the reading frame"
FT /trans= except- (pos:1647..1650; aa:Pro)
FT /note= "this codon has an apparent 1 nucleotide
FT insertion, which alters the reading frame; this
FT insertion is not indicated in the sequence
FT present in the form; sequence listing of the
FT specification."
XX
PN EP939128-A2.
XX
XX 01-SEP-1999.
XX
XX 17-SEP-1990; 99EP-0101746.
XX
XX 15-SEP-1989; 89US-0408045.
XX 21-DEC-1989; 89US-0456142.
XX 17-SEP-1990; 90EP-0310149.
XX
XX (CHIR ) CHIRON CORP.
XX (OYAA/) OYA A.
XX
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA:
PI Miyamura T, Saito I, Weiner AJ:

```

```

XX WFI: 1999-480843/41.
DR P-PSDB; AAY14975.
XX
XX New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
PI infections and development of vaccines
XX
XX Disclosure; Fig 12: 132pp; English.
XX
XX The invention provides two new isolates of hepatitis C virus (HCV), J1
CC and J7. These two isolates comprise nucleotide and amino acid sequences
CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences
CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
CC hybridisation for diagnosis of NANBH infections. They may also be used to
CC screen blood donors, donated blood and blood products for this infection.
CC The isolates may also be used to isolate other naturally occurring
CC variants of the virus. The polypeptides may be used as a vaccine for
CC administration to patients to protect against infection with NANBH. The
CC present sequence represents the nucleotide sequence of HCV-1 ORF.
CC (Updated on 20-MAR-2003 to correct PF field.)
CC (Updated on 20-MAR-2003 to correct PR field.)
XX
SQ Sequence 9133 BP; 1834 A; 2772 C; 2600 G; 1927 T; 0 other;

Alignment Scores:
Pred. No.: 1.74e-77 Length: 9133
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0
DB: 20 Gaps: 0

US-09-965-594-1 (1-182) x AAZ07656 (1-9133)
Qy 1 MetAlaProIleThrAlaTyrAlaGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
Db 3343 CTGGCGCCCATCAGCGGTACGCCAGCAGCAGAGGGGCTCTAGGTGTCATAATCACC 3402
Qy 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
Db 3403 AGCCTAACTGGCGGGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCAACTGCT 3462
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 3463 GCCCAACCTTCTCTGGCAACGTGCATCAATGGGTGTGTGGACTCTACCCAGGGGCG 3522
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 3523 GGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTATCCAGATGTATACCAATGTAGAC 3582
Qy 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 3583 CAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCGCTCATTTGACACCCCTGCCTTGC 3642
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
Db 3643 GGCTCTCGGACCTTACCTGGTGCAGGCGACGCCGATGTCATTCCTGCGCGCGGCGG 3702
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 3703 GGTGATAGCAGGCGGCGCTGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3762
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 3763 GGGGTCTCGCTGTGTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3822
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGlnSerLeuGluThrThrMet 180
Db 3823 ACCCGTGGAGTGGCTAAGCGGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 3882
Qy 181 ArgSer 182
Db 3883 AGGTCC 3888

```

```

RESULT 13
AAQ05956
ID AAQ05956 standard; DNA: 9185 BP.
AC AAQ05956;
XX
XX
XX 25-MAR-2003 (updated)
DT 23-JAN-1991 (first entry)
XX
XX Sense strand of the compiled Hepatitis C virus cDNA sequence.
DE
XX Hepatitis C virus (HCV); antiviral agent: ss.
KW
XX Hepatitis C virus.
OS
XX
XX Key Location/Qualifiers
FH CDS 320..9185
FT FT /*tag= a
FT misc_RNA 1..1667
FT /*tag= b
FT /*note=*epitope within this region is claimed*
FT misc_RNA 8978..9185
FT /*tag= c
FT /*note=*encodes an epitope that is claimed*
XX
XX EP388232-A.
XX
XX PD 19-SEP-1990.
XX
XX PF 16-MAR-1990; 90EP-0302866.
XX
XX PR 17-MAY-1989; 89US-0355002.
XX PR 18-MAR-1989; 89US-0325338.
XX PR 20-APR-1989; 89US-0341334.
XX
XX (CHIR ) CHIRON CORP.
XX
XX PI Houghton M, Choo QL, Kuo G;
XX
XX DR WPI: 1990-284418/38.
XX DR P-PSDB; AAR08124.
XX
XX Hepatitis C virus DNA - used for producing probes,
XX polypeptide(s), antibodies and anti-sense polynucleotide(s) for
XX diagnosis and therapy.
XX
XX PS Disclosure; Fig 17; 83pp; English.
XX
XX CC HCV cDNA libraries were constructed using pooled serum from a
XX chimpanzee with chronic HCV infection. A lambda gtl1 library was
XX screened with probes derived from previously isolated clones. The
XX ORF is derived from the overlapping clones b114a, ag30a, CA205a,
XX CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 26j, 13i,
XX 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
XX 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh. These
XX clones extend the sequence of the HCV genome reported in Ep-318216.
XX The upstream region from nucleotides -319 to -1348 (-1-1667 in this
XX file) is covered by clones b114a, 18g, ag30a, CA205a, CA290a,
XX CA216a, p14a, CA167b, CA156e, CA84a and CA59a; nucleotides
XX 8659-8866 (-8978-9185 in this file) are covered by clones b5a and
XX 16jh.
XX
XX CC See also AAQ05955.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 9185 BP; 1849 A; 2790 C; 2608 G; 1938 T; 0 other;

Alignment Scores:
Pred. No.: 1.75e-77 Length: 9185
Score: 943.00 Matches: 179
Percent Similarity: 100.00% Conservatives: 3
Best Local Similarity: 98.35% Mismatches: 0
Query Match: 98.95% Indels: 0

```

---

```

DB: 11 Gaps: 0
US-09-965-594-1 (1-182) x AAQ05956 (1-9185)
QY 1 MetAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr 20
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3395 CTGGCGCCCATCAGCGCTAGCCAGCAGACAAGGGCCCTCTAGGTGCATATACACC 3454
QY 21 SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAla 40
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3455 AGCCTAACGTGGCGGGGACAAAACCAAGTGGAGGTGAGGTCCAGATTGTCTCAACTGCT 3514
QY 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGlyAla 60
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3515 GCCCAAACTTCCTGGCAACGTGCATCAATGGGGTGTGCTGGACTGTCTACCACGGGGCC 3574
QY 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3575 GGAACGAGGACCATCGCGTCACCAAGGGTCCTGTTCATCCAGATGTATACCAATGTAGAC 3634
QY 81 LysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3635 CAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCTGACCTGC 3694
QY 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3695 GGCTCCTCGGACCTTTACCTGCTCAGAGGACGCGCATGTCTCCGCTGCGCGGCGG 3754
QY 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3755 GGTGATAGAGGGGCGAGCCTGCTGCGCCCGGCGGACGCGGTGGGCATATTTAGGGCCGCGGTGC 3814
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3815 GGGGGTCCGCTGTGTGCGCCCGGCGGACGCGGTGGGCATATTTAGGGCCGCGGTGC 3874
QY 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 3875 ACCCGTGAGTGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGAACCATG 3934
QY 181 ArgSer 182
   ::::::::::
Db 3935 AGGTCC 3940

```

---

```

RESULT 14
AAQ10566
ID AAQ10566 standard; DNA: 9185 BP.
XX
XX AC AAQ10566;
XX
XX DT 25-MAR-2003 (updated)
XX DT 29-APR-1991 (first entry)
XX
XX DE Hepatitis C virus strain 1 DNA.
XX
XX KW Hepatitis C virus; HCV-1; non-A, non-B hepatitis; HCV antigen;
XX viral infections; ss.
XX
XX OS Hepatitis C virus.
XX
XX PN EP414475-A.
XX
XX PD 27-FEB-1991.
XX
XX PF 21-AUG-1990; 90EP-0309120.
XX
XX PR 25-AUG-1989; 89US-0398667.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Weiner AJ, Steimer KS;
XX
XX WPI; 1991-059670/09.

```



```
Qy 41 AlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAla 60
Db 3515 GCCCAAACTTCTGCGAAGTGCATCAATGGGGTGTGIGGACTGCTACCACGGGGCC 3574
Qy 61 GlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnValAsp 80
Db 3575 GGAACGAGGACCATCGCGCTCACCAAGGTCCTGTCTATCCAGATGTATACCAATGTAGAC 3634
Qy 81 LysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThrCys 100
Db 3635 CAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCGGCTCATTTGACACCCCTGCCTTGC 3694
Qy 101 GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 120
Db 3695 GGCTCCTCGGACCTTTACCTGGTCACGAGGACGCCGATGTCATTCCTGTGGCGCGCGS 3754
Qy 121 GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 140
Db 3755 GGTGATAGCAGGGGCGAGCTGTCTCGCCCGCGCCCATTTCTTACTTGAAAGGCTCCTCG 3814
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCys 160
Db 3815 GGGGCTCGGCTGTGTGCCCCGGGGGCGACGCCGTGGGCATATTTAGGGCCGCGGTGTGC 3874
Qy 161 ThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMet 180
Db 3875 ACCCGTGGAGTGGCTAAGCGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACCATG 3934
Qy 181 ArgSer 182
Db 3935 AGGTCC 3940
```

Search completed: August 30, 2003, 19:47:38  
Job time : 188.009 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.01384 Seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-1

Perfect score: 953

Sequence: 1 MAPITAYAOQIRGLIGCIIT.....GVAKAVDFIPVES!FTTMS 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	943	99.0	3011	1 POLG_HCV1	P26664 h genome po
2	927	97.3	3011	1 POLG_HCVH	P27958 h genome po
3	892	93.6	3010	1 POLG_HCVTW	P29846 h genome po
4	891	93.5	3010	1 POLG_HCVBK	P26663 h genome po
5	887	93.1	3010	1 POLG_HCVJA	P26662 h genome po
6	884	92.8	3010	1 POLG_HCVJT	Q00269 h genome po
7	714	74.9	3033	1 POLG_HCVJ8	P26661 h genome po
8	712	74.7	3033	1 POLG_HCVJ6	P26660 h genome po
9	87	9.1	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
10	84	8.8	321	1 RHQA_ARATH	Q9sel7 arabidopsis
11	82	8.6	452	1 AAMP_HUMAN	Q13685 homo sapien
12	82	8.6	485	1 Y136_TREPA	O83172 treponema p
13	80.5	8.4	437	1 DEGL_ARATH	O22609 arabidopsis
14	75	7.9	253	1 CAC3_BOVIN	P05805 bos taurus
15	74.5	7.8	415	1 ZP3_RABIT	P48833 oryctolagus
16	74.5	7.8	776	1 HYPE_AZOVI	P40596 azotobacter
17	74.5	7.8	911	1 TB11_NEIMB	Q09056 neisseria m
18	74	7.8	326	1 PANE_RHIL0	Q987n5 rhizobium l
19	73.5	7.7	263	1 GRAK_MOUSE	O35205 mus musculus
20	73	7.7	730	1 HELS_METMA	Q8prr7 methanosarc
21	72.5	7.6	257	1 GRAM_HUMAN	P51124 homo sapien
22	72	7.6	627	1 CALJ_MOUSE	Q60710 mus musculus
23	72	7.6	1527	1 CAJH_MOUSE	P35061 mus musculus
24	72	7.6	2663	1 CENE_HUMAN	Q02224 homo sapien
25	72	7.6	3491	1 ERV1_SACER	Q03131 saccharopol
26	71.5	7.5	248	1 GRAD_MOUSE	P11033 mus musculus
27	71.5	7.5	323	1 VPRT_SMRVH	P21407 squirrel mo
28	71	7.5	219	1 SPRI_IPOBA	P14715 ipomeia bat
29	71	7.5	336	1 ULL16_EBV	P03221 Epstein-bar
30	71	7.5	529	1 PGL2_RALSO	P20041 raietonia s
31	71	7.5	1180	1 ITA1_RAT	P18614 rattus norv
32	70.5	7.4	264	1 CTRL_HUMAN	P40313 homo sapien
33	70.5	7.4	659	1 VSR2_HEVME	Q03500 hepatitis e

34	70.5	7.4	743	1 TFE3_HUMAN	P19532 homo sapien
35	70	7.3	280	1 VIE3_HCMVT	P06434 human cytom
36	70	7.3	410	1 VIE2_HCMVT	P06435 human cytom
37	70	7.3	478	1 MM03_RABIT	P28863 oryctolagus
38	70	7.3	730	1 HELS_METAC	Q8t139 methanosarc
39	69.5	7.3	915	1 TBPI_NEIGO	Q01996 neisseria g
40	69	7.2	355	1 CWG2_SCHPO	P32434 schizosacch
41	69	7.2	637	1 DNAK_BRUME	Q8ye76 bruceella me
42	69	7.2	1136	1 C4BA_BACTI	P05519 bacillus th
43	69	7.2	1180	1 C4AA_BACTI	P16480 bacillus th
44	69	7.2	4391	1 PGBM_HUMAN	P98160 homo sapien
45	68.5	7.2	255	1 CATG_HUMAN	P08311 homo sapien

#### ALIGNMENTS

##### RESULT 1

POLG\_HCV1 STANDARD; PRT; 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DE Hepatitis C virus (isolate 1) (HCV).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
OC NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;  
RA \*Genetic organization and diversity of the hepatitis C virus.\*;  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC EMBL; M62321; AAA45676.1; -  
DR PIR; A39166; GNMV03  
DR PDB; 1AIV; 16-FEB-99.  
DR PDB; 1HEI; 25-NOV-98.  
DR MEROPS; S29.001; -  
DR MEROPS; U39.001; -  
DR InterPro; IPR001410; DEAD  
DR InterPro; IPR002522; HCV\_capsid.



```

DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007084; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 MATRIX PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;

Query Match 99.0%; Score 943; DB 1; Length 3011.
Best Local Similarity 98.48; Pred. No. 3.4e-82.
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPITAYAQOTRGLGCIITSITGRDNQVEGEVQIVSTAOTFLATCINGCVTVYHGA 60

```

```

Db 1026 LAPITAYAQOTRGLGCIITSITGRDNQVEGEVQIVSTAOTFLATCINGCVTVYHGA 1085
Oy 61 GTRTIASPKGPVITOMYTNVDKDLVGNPAPQGGSRSLTPCTCGSSDLYLTRHADVTPVRRR 120
Db 1086 GTRTIASPKGPVITOMYTNVDQDLVGNPAPQGGSRSLTPCTCGSSDLYLTRHADVTPVRRR 1145
Oy 121 GDSRGLLSRPISYLUKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
Db 1146 GDSRGLLSRPISYLUKSGGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
Oy 181 RS 182
Db 1206 RS 1207

RESULT 2
POLG_HCVH STANDARD; PRT: 3011 AA.
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DI 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.46)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE-92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN [2]
X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE-97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE-98154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

```

CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M67463; AAA45534.1; -  
CC PIR: A36814; GNVVCH.  
CC PDB: 1HEJ; 25-NOV-98.  
CC PDB: 1A1V; 16-FEB-99.  
CC PDB: 1A1R; 17-JUN-98.  
CC MEROPS: S29.001; -  
CC MEROPS: U39.001; -  
CC TRANSFAC: T04155; -  
CC InterPro: IPR001410; DEAD.  
CC InterPro: IPR002522; HCV\_capsid.  
CC InterPro: IPR002521; HCV\_core.  
CC InterPro: IPR002519; HCV\_env.  
CC InterPro: IPR002531; HCV\_NS1.  
CC InterPro: IPR002518; HCV\_NS2.  
CC InterPro: IPR004109; HCV\_NS3.  
CC InterPro: IPR000745; HCV\_NS4a.  
CC InterPro: IPR001490; HCV\_NS4b.  
CC InterPro: IPR002868; HCV\_NS5a.  
CC InterPro: IPR002166; HCV\_RdRP.  
CC InterPro: IPR001650; Helicase\_C.  
CC InterPro: IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro: IPR007094; RNA\_pol\_PSVir.  
CC Pfam: PF01543; HCV\_capsid\_1.  
CC Pfam: PF01542; HCV\_core\_1.  
CC Pfam: PF01539; HCV\_env\_1.  
CC Pfam: PF01560; HCV\_NS1\_1.  
CC Pfam: PF01538; HCV\_NS2\_1.  
CC Pfam: PF02907; HCV\_NS3\_1.  
CC Pfam: PF01006; HCV\_NS4a\_1.  
CC Pfam: PF01001; HCV\_NS4b\_1.  
CC Pfam: PF01506; HCV\_NS5a\_1.  
CC Pfam: PF00271; helicase\_C\_1.  
CC Pfam: PF00998; Viral\_RdRP\_1.  
CC ProDom: PD186062; HCV\_NS1\_1.  
CC SMART: SM00487; DEXDC; 1.  
CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
CC 3D-structure.  
CC INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.  
FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.  
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.  
FT CHAIN 747 809 PROTEIN P7.  
FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.  
FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.  
FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.  
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.  
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.  
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.  
FT CHAIN 3011 369 POTENTIAL.  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
FT SITE 1316 1319 DECH\_BOX.  
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT 305 305 CARBOHYD  
FT 417 417 CARBOHYD  
FT 423 423 CARBOHYD  
FT 430 430 CARBOHYD  
FT 430 430 CARBOHYD  
FT 448 448 CARBOHYD  
FT 448 448 CARBOHYD  
FT 476 476 CARBOHYD  
FT 532 532 CARBOHYD  
FT 540 540 CARBOHYD  
FT 556 556 CARBOHYD  
FT 576 576 CARBOHYD  
FT 623 623 CARBOHYD  
FT 645 645 CARBOHYD  
FT 1224 1224 STRAND  
FT 1232 1232 TURN  
FT 1236 1236 TURN  
FT 1238 1238 TURN  
FT 1246 1246 HELIX  
FT 1247 1247 TURN  
FT 1248 1248 TURN  
FT 1251 1251 STRAND  
FT 1255 1255 STRAND  
FT 1258 1258 HELIX  
FT 1271 1271 TURN  
FT 1272 1272 TURN  
FT 1277 1277 STRAND  
FT 1280 1280 TURN  
FT 1281 1281 TURN  
FT 1283 1283 STRAND  
FT 1291 1291 STRAND  
FT 1295 1295 STRAND  
FT 1301 1301 HELIX  
FT 1303 1303 TURN  
FT 1312 1312 STRAND  
FT 1316 1316 TURN  
FT 1319 1319 TURN  
FT 1323 1323 HELIX  
FT 1336 1336 TURN  
FT 1340 1340 TURN  
FT 1343 1343 STRAND  
FT 1352 1352 TURN  
FT 1353 1353 TURN  
FT 1361 1361 TURN  
FT 1362 1362 STRAND  
FT 1368 1368 STRAND  
FT 1373 1373 STRAND  
FT 1375 1375 TURN  
FT 1376 1376 TURN  
FT 1382 1382 STRAND  
FT 1389 1389 STRAND  
FT 1393 1393 STRAND  
FT 1397 1397 HELIX  
FT 1409 1409 TURN  
FT 1410 1411 TURN  
FT 1414 1417 TURN  
FT 1419 1420 TURN  
FT 1432 1436 STRAND  
FT 1438 1439 TURN  
FT 1450 1453 STRAND  
FT 1456 1463 STRAND  
FT 1471 1478 STRAND  
FT 1480 1480 STRAND  
FT 1481 1488 HELIX  
FT 1489 1490 TURN  
FT 1497 1501 STRAND  
FT 1507 1507 STRAND  
FT 1511 1511 STRAND  
FT 1514 1527 HELIX  
FT 1532 1544 HELIX  
FT 1550 1550 STRAND  
FT 1555 1564 HELIX  
FT 1570 1578 HELIX  
FT 1579 1580 TURN  
FT 1584 1597 HELIX  
FT 1598 1598 TURN  
FT 1606 1611 HELIX  
FT 1614 1618 TURN  
FT 1622 1623 STRAND  
FT 1627 1627 STRAND  
FT 1635 1636 STRAND  
FT 1640 1652 HELIX  
FT 3011 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;  
SQ SEQUENCE  
Query Match 97.3%; Score 927; DB 1; Length 3011;  
Best Local Similarity 96.2%; Pred. No. 1.2e-80;

Matches 175; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPITAYAQTRGLGCGIITSLTRGDKNOVEGVQIVSTAQTFLATCNGVCWTYYHGA 60  
 :|||||  
 Db 1026 LAPITAYAQTRGLGCGIITSLTRGDKNOVEGVQIVSTATOTFLATCNGVCWTYYHGA 1085  
 :|||||  
 Qy 61 GTRTIAAPKGVQMTYNTVDKLVGPAPOGSRSLTPCTCGSSDYLVTTRHADVTPVRRR 120  
 :|||||  
 Db 1086 GTRTIAAPKGVQMTYNTVDKLVGPAPOGSRSLTPCTCGSSDYLVTTRHADVTPVRRR 1145  
 :|||||  
 Qy 121 GDSRGLSPRPISYLYKSGSGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETIM 180  
 :|||||  
 Db 1146 GDSRGLSPRPISYLYKSGSGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETIM 1205  
 :|||||  
 Qy 181 RS 182  
 :||  
 Db 1206 RS 1207  
 :||

RESULT 3  
 POLG\_HCVTW STANDARD; PRT; 3010 AA.

AC P29846;  
 DT 01-APR-1993 (Rel. 25, last sequence update)  
 DT 15-SEP-2003 (Rel. 42, last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID-31645;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92230206; PubMed-1314449;  
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RT "The Taiwanese hepatitis C virus genome: sequence determination and  
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S23.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M84754; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A40244; GNVVTV.  
 DR PDB; 1N64; 25-FEB-03.  
 DR PDB; 1NS3; 08-APR-98.  
 DR MEROPS; S29.001; -;  
 DR InterPro; IPR001410; DFAD.

DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR004109; HCV\_NS3.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF03998; Viral\_RdRP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure. 1  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1815  
 FT CHAIN 1816 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDPE215 CRC64;

Query Match 93.6%; Score 892; DB 1; Length 3010;  
 Best Local Similarity 90.1%; Pred. No. 2,7e-77;  
 Matches 164; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAPITAYAOOTGGLCCITTSITGRKKNQVEGVCQVSTAAQTFLATCINGCVCTVYHGA 60  
 DB 1026 LAPITAYAOOTGGLCCITTSITGRKKNQVEGVCQVSTAAQTFLATCINGCVCTVYHGA 1085  
 QY 61 GTTITASPKGPVIOMYNNVDLVGMPAPQGSRSRLTPTCGSSDLYLVTRHADYIPVRRR 120  
 DB 1086 GSKTLAGPKGPITOMYNNVDLVGMPAPQGSRSRLTPTCGSSDLYLVTRHADYIPVRRR 1145  
 QY 121 GDSRGLSPRISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 DB 1146 GDSRGLSPRISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESMETTM 1205  
 QY 181 RS 182  
 DB 1206 RS 1207

RESULT 4  
 ID POLG\_HCVBK STANDARD; PRT: 3010 AA.  
 AC P26663;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48);  
 OS Hepatitis C virus (isolate BK) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11105;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91140698; PubMed=18477440;  
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,  
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;  
 RT "Structure and organization of the hepatitis C virus genome isolated  
 RT from human carriers.";  
 RL J. Virol. 65:1105-1113(1991).  
 RN [2]  
 RP SEQUENCE OF 1487-1500.  
 RX MEDLINE=96235224; PubMed=8647104;  
 RA Borowski P., Helland M., Oehlmann K., Becker B., Kornteky L.;  
 RT "Non-structural protein 3 of hepatitis C virus inhibits  
 RT phosphorylation mediated by cAMP-dependent protein kinase.";  
 RL Eur. J. Biochem. 237:611-618(1996).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.  
 RX MEDLINE=97015088; PubMed=8861916;  
 RA Love R.A., Parge H.E., Wickorsham J.A., Hostomsky Z., Habuka N.,  
 RA Moomaw E.W., Adachi T., Hostomsky Z.;  
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a  
 RT trypsin-like fold and a structural zinc binding site.";  
 RL Cell 87:331-342(1996).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.  
 RX MEDLINE=98227846; PubMed=9568891;  
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,  
 RA Steinkuehler C., Tonel L., de Francesco R., Kuo L.C., Chen Z.;  
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C  
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";  
 RL Protein Sci. 7:837-847(1998).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

(RNA1(N)).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M58335; AAA72945.1; -  
 CC PIR: A38465; GHWYTC.  
 CC PDB: 1A1Q; 25-MAR-98.  
 CC PDB: 1JXP; 14-JAN-98.  
 CC PDB: 1NS3; 08-APR-98.  
 CC PDB: 1C2P; 15-NOV-00.  
 CC PDB: 1CSJ; 08-NOV-99.  
 CC PDB: 1GX5; 09-APR-02.  
 CC PDB: 1GX6; 10-APR-02.  
 CC PDB: 1QUV; 26-JUN-00.  
 CC PDB: 8OHM; 20-APR-99.  
 CC MEROPS: S29.001; -  
 CC MEROPS: U39.001; -  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV\_capsid.  
 CC InterPro: IPR002521; HCV\_Core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC InterPro: IPR002531; HCV\_NS1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR004109; HCV\_NS3.  
 CC InterPro: IPR000745; HCV\_NS4a.  
 CC InterPro: IPR001490; HCV\_NS4b.  
 CC InterPro: IPR002868; HCV\_NS5a.  
 CC InterPro: IPR002166; HCV\_RdRP.  
 CC InterPro: IPR007095; RNA\_pol\_PS.  
 CC InterPro: IPR007094; RNA\_pol\_PSVir.  
 CC Pfam: PF01543; HCV\_capsid; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01539; HCV\_env; 1.  
 CC Pfam: PF01560; HCV\_NS1; 1.  
 CC Pfam: PF01538; HCV\_NS2; 1.  
 CC Pfam: PF02907; HCV\_NS3; 1.  
 CC Pfam: PF01006; HCV\_NS4a; 1.  
 CC Pfam: PF01001; HCV\_NS4b; 1.  
 CC Pfam: PF01506; HCV\_NS5a; 1.  
 CC Pfam: PF00998; Viral\_RdRP; 1.  
 CC ProDom: PD186062; HCV\_NS1; 1.  
 CC SMART: SM00487; DEXDC; 1.  
 CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 CC 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 3010 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM.  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).

```

FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 1035 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT STRAND 1101 1103
FT TURN 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1680 1688
SQ SEQUENCE 3010 AA: F8422D5ECCDFD9C CRC64:

```

```

Query Match 93.5% Score 891; DB 1; Length 3010;
Best Local Similarity 89.0%; Pred. No. 3.4e-77;
Matches 162; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTGGLGCIITSLTGDRKNQVGEVOIVSTAQTATCINCVCVTVYHGA 60
DB 1026 LAPITAYSQOTRGLGCIITSLTGDRKNQVGEVOIVSTAQTATCINCVCVTVYHGA 1085

QY 61 GRTIASPGPVITOMYTNVDKLVGWPAPOGSRSLTPCTCGSSDLVLYTRHADVIPVRRR 120
DB 1086 GSKTLPAGKPTOMYTNVDQDLVGVKPPGARSILTPCTCGSSDLVLYTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYIKGSGGGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVPSLEITM 180
DB 1146 GDSRGSLLSPRPVSYLKSGSGGLPCPGHAVGIFRAAVCTRGVAKAVDFVPSMETTM 1205

QY 181 RS 182
  11

```

```

Db 1206 RS 1207

RESULT 5
POLG_HCVJA STANDARD; PRT; 3010 AA.
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26652;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
  Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
  (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
  (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
  (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
  NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
  NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
  Hepacivirus.
OX NCBI_taxid=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91088550; Pubmed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
  Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
  Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE-91192160; Pubmed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
  Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
  HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
  precursor polyprotein, commonly with Asp or Glu in the P6
  position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
  (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
  PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by, and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC EMBL: D90208; BAA14233.1; -.
DR PIR: A39253; GNVVCJ.
DR NSSP: P26663; LJAP.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002531; HCV env.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR004109; HCV NS3.
DR InterPro: IPR00745; HCV NS4.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.

```











DR Pfam: PF02907; HCV\_NS3: 1.  
 DR Pfam: PF01006; HCV\_NS4a: 1.  
 DR Pfam: PF01001; HCV\_NS4b: 1.  
 DR Pfam: PF01506; HCV\_NS5a: 1.  
 DR Pfam: PF00271; helicase\_C: 1.  
 DR Pfam: PF00998; Viral\_RdRP: 1.  
 DR ProDom: PD186062; HCV\_NS1: 1.  
 DR SMART: SM00487; DEXDC: 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT CHAIN 3033 369  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA: 329165 MW: 74.7kDa; Score 712; DB 1; Length 3033;  
 Best Local Similarity 69.8%; Pred. No. 5.3e-60;  
 Matches 127; Conservative 29; Mismatches 26; Indels 0; Gaps 0;  
 QY 1 MAPITAYAQQTGLGCIITSLTGRDNQVEGVQIVSTAQTFLATCINGVCWTVYHGA 60  
 DB 1030 LAPITAYAQQTGLGIIITVYMTGRDKTEQAGIQVLSTVTSFGLTITSGVLWTVYHGA 1089  
 QY 61 GTRTASPKGPVQMTYNTVDKLVGAPAGSRLTPTCTGSSDLYLVTRHADVTPVRRR 120  
 DB 1090 GNKTLAGSRGVTQMTSSAGDLVGMSPGPTKSLPCTCGAVDLYLVTRHADVTPVRRR 1149  
 QY 121 GDSRGLSPRPISYLGKSSGGPLLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLEITM 180  
 DB 1150 GDKRGALLSPRLSTLKGSSGGPVLCPRGHAGVIFRAAVCTRGVAKAVDFIPVESLEITM 1209  
 QY 181 RS 182  
 DB 1210 RS 1211  
 RESULT 9  
 PAAD\_PSEAE STANDARD; PRT: 209 AA.  
 AC Q9HX08;

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).  
 GN PA4019.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 [1]  
 SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizioch S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.K., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT \*Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.;  
 RL Nature 406:959-964(2000).  
 CC -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /  
 CC PHENYLACETYLIC ACID DECARBOXYLASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE004818; AAG07406.1;  
 CC PIR: H83144; H83144.  
 DR InterPro: IPR003182; Flavoprotein.  
 DR Pfam: PF02441; Flavoprotein; 1  
 KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.  
 SQ SEQUENCE 209 AA: 22367 MW: 01FD08ICC495D3F6 CRC64;  
 Query Match 9.1%; Score 87; DB 1; Length 209;  
 Best Local Similarity 26.2%; Pred. No. 0.28;  
 Matches 55; Conservative 18; Mismatches 61; Indels 76; Gaps 12;  
 QY 8 AQQTGILGCIITSLTGRDNQVEGVQ-IVSTAQTFLATCINGVCWTVYHGA GTRTIA 66  
 DB 17 AQYGLRLDCLV-----QEEREVHFLISKAAQLVMAT-----ETDVA 53  
 QY 67 SPKGP-----VIQMTNVDKLDLVGAPAGSRLTPTCTGSSDLYLVTRHADVTPVRRR 105  
 DB 54 LPAPQAMQAFLEYCGAAGQIVFQND-----HWAPPAGSSAPNAPWICPSTGT 108  
 QY 106 -----YLVTRHADVTPVRRRGRSGLSLSPR--PIS-----YLGSSGGPLLCRA 148  
 DB 109 SAVATGACNNLIERAADVALKER----RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA 164  
 QY 149 GHAVGIETRAAVCTRGVAKAVDFIPVESLET 178  
 DB 165 --APGFTHQ----POSVEDLVDFVVARILNT 189  
 RESULT 10  
 ID HHOA\_ARATH  
 AC Q9SEL7; Q9507;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).  
 GN HHOA OR AT4G18370 OR F28J12.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF114386; AAF24060.1; -  
EMBL: AL021710; CAA16717.1; ALT\_SEQ.  
EMBL: AL161548; CAB78839.1; ALT\_SEQ.  
MEROPS: S01.279; -  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR PRINTS: PR00834; PROTEASES2C.  
KW Hydrolyase; Serine protease; Chloroplast; Thylakoid; Transit peptide.  
FT TRANSIT 1 26 CHLOROPLAST (POTENTIAL).  
FT CHAIN 27 71 THYLAKOID.  
FT DOMAIN 77 87 POLY-GLU.  
FT ACT\_SITE 145 145 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 186 186 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 264 264 CHARGE RELAY SYSTEM (POTENTIAL).  
FT CONFLICT 40 40 R -> G (IN REF. 1).  
SQ SEQUENCE 321 AA: 34691 MW: 68081E08D27A7A7 CRC64:  
Query Match Score 84; DB 1; Length 321;  
Best Local Similarity 22.8%; Pred. No. 0.88;  
Matches 48; Conservative 26; Mismatches 60; Indels 82; Gaps 11;  
QY 22 LTGRKQKQVEGVQIVSTAQTFLATCINGVW-----TVYH----- 58  
DB 117 LTDENGKIEGTG-----SGFVMDKLGHVNYHVIKATDQFLQRCX 161  
QY 59 -----GACTRTIASPKPVIQMYTVNDRKDLGWPAQGSRLTPTCGSSDLYLVTRHAD 113  
DB 162 VSLVDKAGTR--FSKEGKIVGL--DPDNDLAVLKLTETEGRELNPVLTGNSDLRWGSCF 217  
QY 114 VIPVRGRDSRG-----SLLSPRISYLYK-----GSSGGPILCPA 148  
DB 218 AI-----GNPYGENTLTIGVVGSLGRIPSPNGKSISEATOTADINSNGSGPLDSDY 272  
QY 149 GHAVGIFRAAVCTR---GVAKAVDF-IPVESLETTM 180  
DB 273 GHTIGV-NTATFTKSGMSSGVNFAIDIVTVRV 307  
STANDARD: PRT: 452 AA.  
RESULT 11  
AAMP\_HUMAN  
ID AAMP\_HUMAN STANDARD: PRT: 452 AA.  
AC Q13685;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Angio-associated migratory cell protein.  
CN AAMP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=95262124; PubMed=7743515;  
RA Beckner M.E., Krutzsch H.C., Stracke M.L., Williams S.T.,  
RA Gallardo J.A., Liotta L.A.;  
RT "Identification of a new immunoglobulin superfamily protein expressed  
RT in blood vessels with a heparin-binding consensus sequence.";  
RL Cancer Res. 55:2140-2149 (1995).  
CC -!- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED  
CC IN ENDOTHELIAL CELLS. CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED  
CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC

eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
NCBI\_TaxID=3702;  
[1]  
SEQUENCE FROM N.A.  
Lensch M.H.A., Sokolenko A., Herrmann R.G.;  
\*Identification and characterization of the chloroplast HhoA protease,  
a homolog to the bacterial periplasmic protease HhoA.\*  
Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.  
STRAIN=cv. Columbia;  
MEDLINE=20083488; PubMed=10617198;  
BA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
BA Pohl T., Duesterhoeft A., Stickens W., Entian K.-D., Terryn N.,  
BA Harris B., Ansurge W., Brandt P., Grivell L., Rieger M.,  
BA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
BA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
BA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
BA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
BA Langham S.-A., McCullagh B., Billham L., Robben J.,  
BA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
BA Braeken M., Weltjens I., Voet M.B., Bastiaens I., Aert R., Deleor E.,  
BA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,  
BA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,  
BA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
BA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
BA De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R.,  
BA Van Montagu M., Rogers J., Cronin A., Ouail M., Bray-Allen S.,  
BA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Hayes R.,  
BA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,  
BA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehnert T.-H.,  
BA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
BA Gabel C., Fuchs M., Farmann B., Graendath K., Dauner D., Herzl A.,  
BA Neumann S., Argüioy A., Vitale D., Liquori R., Piravandi E.,  
BA Massenot O., Quicley F., Clabaud G., Muendlein A., Felber R.,  
BA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
BA Cherdorf F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
BA Gibbons T., Weber N., Vandenbol M., Barges M., Torol J., Torres A.,  
BA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacor D., Jesse T.,  
BA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,  
BA Frishman D., Haase D., Lemcke K., Meves H.-W., Stocker S.,  
BA Zaccaria P., Bevan M., Wilson K.K., de la Hasstide M., Habermann K.,  
BA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel H.,  
BA Sekhon M., Murray J., Smet P., Cordes M., Abu-Frieden J.,  
BA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
BA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
BA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
BA Kramer J., Fulton L., Mardis E., Dancie M., Pepin K., Hillier L.,  
BA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
BA Du H., Ali J., Bergthoff A., Jones K., Drone K., Cotton M., Joshi C.,  
BA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
BA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
BA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
BA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
BA Chen E., Marra M., Martienssen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
RT thaliana".  
RT Nature 402:769-777 (1999).  
RN [3]  
RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.  
RA Schubert M., Peterson U., Funk C., Haas H., Schroeder W.P.,  
RA Kieselbach T.;  
RT "The chloroplast lumen from Arabidopsis thaliana".  
RL Submitted (JUL-2001) to the SWISS-PROT data bank.  
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 52C.  
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
CC gene model prediction. AT4G18370 and AT4G18375 were originally  
CC fused into a single gene.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; M95627; AAA68889.1; -  
 CC PIR; I39383; I39383.  
 CC Genew; HGNC:18; AAMP.  
 CC MIM; 603488; -  
 CC GO; GO:0008201; F:heparin binding activity; TAS.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF00400; WD40; 8.  
 CC SMART; SM00320; WD40; 8.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 CC PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 CC PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 CC Repeat; WD repeat.  
 CC KW Repeat; WD repeat.  
 CC FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).  
 CC FT CHAIN 71 77 POLY-GLU.  
 CC FT REPEAT 107 138 WD 1.  
 CC FT REPEAT 150 180 WD 2.  
 CC FT REPEAT 190 220 WD 3.  
 CC FT REPEAT 231 261 WD 4.  
 CC FT REPEAT 276 306 WD 5.  
 CC FT REPEAT 333 363 WD 6.  
 CC FT REPEAT 374 404 WD 7.  
 CC FT REPEAT 416 446 WD 8.  
 CC SQ SEQUENCE 452 AA; 49015 MW; DA1413D25EB236C0 CRC64;

Query Match 8.6%; Score 82; DB 1; Length 452;  
 Best Local Similarity 25.3%; Pred. No. 2;  
 Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

OY 54 WTIVYHAGTRTIAIPKGVQIMTVNDKLVGPAPGSRSL-----TPCTCGSSDLYLV 108  
 DB 197 WNEWH-----PRAPVLLAGT-ADGNTWMKVPNGDCKTQGGNCPATCG----- 240  
 OY 109 TRHADVIVPVR-----GDSRGS-----LLSPRPISYKLGSSG--GPILCPA----- 148  
 DB 241 -----VLPDGRKRAVGVYEDGTIRIWLKQSPHVLKGTGEGHQLTCVAAHQDGSLLT 295  
 OY 149 -----GHAVGIFR-----RAVCTRGVAKAVDFIPVESL 176  
 DB 296 GSVDCQAKLVSAATTKVGVGVFRPETVASQFSLGRGESESNVSLSL 341

RESULT 12  
 Y136\_TREPA  
 ID Y136\_TREPA STANDARD; PRT; 485 AA.  
 AC O83172;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical lipoprotein TP0136 precursor.  
 GN TP0136.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RC STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9655876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RC "Complete genome sequence of Treponema pallidum, the syphilis

RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 CC (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AE001199; AAC65137.1; ALT\_INIT.  
 CC TIGR; TP0136; -  
 CC KW Hypothetical protein; Lipoprotein; Membrane; Signal;  
 KW Complete proteome.  
 CC FT SIGNAL 1 23 POTENTIAL.  
 CC FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.  
 CC FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).  
 CC FT DOMAIN 164 178 GLY/SER-RICH.  
 CC FT DOMAIN 196 210 GLY/SER-RICH.  
 CC FT DOMAIN 253 267 GLY/SER-RICH.  
 CC FT DOMAIN 318 327 POLY-SER.  
 CC FT DOMAIN 444 447 POLY-SER.  
 CC SQ SEQUENCE 485 AA; 48984 MW; C7A4CEDC7DC5CED CRC64;

Query Match 8.6%; Score 82; DB 1; Length 485;  
 Best Local Similarity 24.2%; Pred. No. 2;  
 Matches 44; Conservative 13; Mismatches 65; Indels 60; Gaps 8;

OY 23 TGRDKNOVEGEVQIVSTAQTFLATCI--NGVCWTVYHGAG---TRTIASPKGPVQMVT 77  
 DB 86 TDSK-----KMSIATDGTFTVLACVPGTGKVCNAGAGSSSTGTASPSTETCSQHA 140  
 OY 78 NVDRKLVG-----WPAQGSRSILTPCTC-----GSSDLYLVTRHADVIP-----VR 118  
 DB 141 T-----LVGGTSKPLVPGGTGNGNCGCGGGGGSSSSSSSIHILWLVPGGTGNGNCG 196  
 OY 119 RRGDSRGSLLSPRISYK-----GSSGGPLLCFAGHA 151  
 DB 197 CGGGGGSSSSSSSIHILKVENTQFLDMGEGYVTTKHLTKNGSSSAGPAQCPCGGG 256  
 OY 152 VG 153  
 DB 257 GG 258

RESULT 13  
 DEGLARATH  
 ID DEGLARATH STANDARD; PRT; 437 AA.  
 AC O22609; Q9LKH5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).  
 GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=98175982; PubMed=9507020;  
 RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;  
 RT "Identification and characterization of Degp, a serine protease  
 RT associated with the luminal side of the thylakoid membrane.";  
 RL J. Biol. Chem. 273:7094-7098(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;

Db 210 PK--NKLRPPIPVGVSAADLLVGQKFAIGNFGDLHTLTGTGVISGLRRREIS--SAATGRPI 265

QY 134 SYL-----KGSSGGPILCPAGHVAAGTCAVAKADF-IPVESL 176  
:  
:::||||| :|||  
:

Dd 266 QDVQTDAAINFCNSGGFLLOSSGTLIGINTALYSFSGASSGVGFSPVDVT 317  
:  
:::||||| :|||  
:

RESULT 14

CAC3\_BOVIN STANDARD; PRT; 253 AA.

ID CAC3\_BOVIN ID AC P05805;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 15-DIC-1998 (Rel. 37, Last sequence update)  
DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Proproteinase E precursor [Procarboxypeptidase A complex component III] (Procarboxypeptidase A-S6 subunit III) (PROCPA-S6 III).  
DE OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
OC Bovidae; Bovinae; Bos.  
OC NCBI\_TaxId=9913;

RX [1]  
RN SEQUENCE OF 1-253.  
RX MEDLINE=91099520; PubMed=2269366;  
RA Pascual R., Vendrell J., Aviles F.X., Bonicel J., Wicker C., Puigserver A.;  
RA \*Autolysis of proproteinase E in bovine procarboxypeptidase A ternary complex gives rise to subunit III.\*;  
FEBS Lett. 277:37-41(1990).

[2]  
RN SEQUENCE OF 14-253, AND DISULFIDE BONDS.  
RX MEDLINE=86220198; PubMed=3519215;  
RA Venot N., Sciaky M., Puigserver A., Desnuelle P., Laurent G.;  
RX \*Amino acid sequence and disulfide bridges of subunit III, a defective endopeptidase present in the bovine pancreatic 6 S procarboxypeptidase A complex.\*;  
Eur. J. Biochem. 157:91-99(1986).

[3]  
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
RX MEDLINE=9422022; PubMed=8168476;  
RA Pignol D., Gaboriaud C., Michon T., Kerfelec B., Chapus C., Fontecilla-Camps J.C.;

RA "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a highly structured truncated zymogen E.";  
EMBO J. 13:1763-1771(1994).

CC -I- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM TO HAVE A PROPEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC ENVIRONMENT OF THE RUMINANT DUODENUM.

CC -I- SUBUNIT: HETEROTRIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND CHYMOTRYPSINOGEN C.

CC -I- SUBCELLULAR LOCATION: Extracellular.

CC -I- TISSUE SPECIFICITY: Pancreas.

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

DR PDB; IFON; 14-OCT-96.  
DR PDB; LPYT; 27-JAN-97.  
DR MEROPS; S01.983; ...

DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SK00020; Tryp\_SPC; 1.  
DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.

KW Serine protease homolog; Pancreas; Digestion; 3D-structure.  
FT PROPEP 1 11 ACTIVATION PEPTIDE.  
FT CHAIN 12 253 PROTEINASE E.  
FT DISULFD 41 57  
FT DISULFD 100 103  
FT DISULFD 140 206  
FT DISULFD 171 187  
FT DISULFD 196 227

```
FT STRAND 23 33
FT TURN 34 35
FT STRAND 36 41
FT STRAND 44 47
FT TURN 48 49
FT STRAND 50 53
FT HELIX 55 57
FT TURN 60 61
FT STRAND 64 72
FT TURN 73 74
FT STRAND 75 84
FT TURN 87 88
FT STRAND 90 92
FT TURN 94 95
FT TURN 98 99
FT HELIX 101 103
FT STRAND 108 111
FT TURN 119 120
FT STRAND 126 126
FT TURN 130 131
FT TURN 136 137
FT STRAND 139 144
FT TURN 146 147
FT STRAND 157 157
FT STRAND 159 166
FT HELIX 168 171
FT TURN 172 172
FT TURN 174 177
FT HELIX 178 180
FT TURN 183 184
FT STRAND 185 188
FT TURN 196 197
FT TURN 200 201
FT STRAND 203 207
FT TURN 209 210
FT STRAND 213 221
FT STRAND 223 223
FT TURN 224 225
FT STRAND 226 226
FT TURN 230 231
FT STRAND 234 238
FT HELIX 239 241
FT HELIX 243 253
SQ SEQUENCE 253 AA; 27337 MW; 24663724D8AE409C CRC64;

Query Match 7.98; Score 75; DB 1; Length 253;
Best Local Similarity 22.78; Pred. No. 4.9;
Matches 46; Conservative 23; Mismatches 72; Indels 62; Gaps 10;

QY 17 CIITSLTGR-----DKNOVEGEVOIVS-TAQTFLATCINGCYWTVYHCAGTRTIA---66
DB 57 CISTRTYQVVLGYEYDSVLEGESEQVPIINAGDLFVHPLWNSVCVACGNDIALVKLSRSA 116
QY 67 -----SPKGPVI-----QMTNVDKOLVGPAPQG-SRSLTPCT---99
DB 117 QLGDVKVOLANLPPAGDILPNEAPCYISGWGNLYT-----GGPLPKLQALLPVVDYE 169
QY 100 -CGSSDLYLVTRHADVPVRRGDSRGLSPRPISYLVKSSGGPLCPAG-----HAV 152
DB 170 HCSQMDWMGITVKKIM--VCAGSDTR-----SGCNGDSGGLNCPAAGDSWQVHGV 218
QY 153 GIFRAAVCTRGVAKAVDFIPVES 175
DB 219 TSFVSATGCTIKKPTVTRVSA 241

RESULT 15
ID 2P3_RABIT STANDARD; PRT: 415 AA.
AC P48833;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
```

```
DE Zona pellucida sperm-binding protein 3 precursor (Zona pellucida
DE glycoprotein ZP3) (Sperm receptor) (Zona pellucida protein C)
DE (Fragment).
GN ZP3 OR ZPC.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95143578; PubMed=7841460;
RA Harris J.D., Hibler D.W., Fontenot G.K., Hsu K.T., Yurewicz E.C.,
RA Sacco A.G.;
RT "Cloning and characterization of zona pellucida genes and cDNAs from
RT a variety of mammalian species: the ZPA, ZPB and ZPC gene families.";
RL DNA Seq. 4:361-393(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05782; AAA74392.1; -.
DR PIR: S70401; S70401.
DR InterPro: IPR001507; Endoglin/CD105.
DR Pfam: PF00100; zona_pellucida; 1.
DR SMART: SM00241; ZP; 1.
DR PROSITE: PS00682; ZP_DOMAIN; 1.
KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW Extracellular matrix; Multigene family.
FT NON_TER 1
FT SIGNAL <1 18 POTENTIAL.
FT CHAIN 19 415 ZONA PELLUCIDA SPERM-BINDING PROTEIN 3.
FT DOMAIN 19 378 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 379 399 POTENTIAL.
FT DOMAIN 400 415 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 41 301 ZP.
SQ SEQUENCE 415 AA; 44987 MW; 77396CF1BA3F5CB CRC64;

Query Match 7.88; Score 74.5; DB 1; Length 415;
Best Local Similarity 26.78; Pred. No. 9.5;
Matches 31; Conservative 17; Mismatches 53; Indels 15; Gaps 5;

QY 55 TVYHCAGTRTITASPKGP-VIQMTNVDKOLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 113
DB 271 TVYITCHLVTPAQAPDRLNKACSFNOSSSWAPVEGSADICEC-CGNGDCDLIAGS--- 327
QY 114 VIPVRRGDSRGLSPRPISYLVKSSGGPL--LCPAGHAVGIFRAAVCTRGVAKA 167
DB 328 --PKNQNHAAARSLRRSRHVTREADVTVGLFLGKAGDPAG-----TEGLASA 374

Search completed: August 30, 2003, 19:13:43
Job time : 13.0138 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model:

Run On: August 30, 2003, 19:00:22 ; Search Time 34.7298 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-1  
Perfect score: 953  
Sequence: 1 MAPITAYAQOTRGLLCIIT.....GVAKAVDFIPVESLETTMRS 182

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : STREMBL_23.*
1: sp_arched.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviro.*
16: sp_bacteriap.*
17: sp_archaeap.*
```

## SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	945	99.2	181	12	Q91RR8	Q91RR8	hepatitis c
2	945	99.2	181	12	Q91RT5	Q91RT5	hepatitis c
3	943	99.0	2436	12	Q81756	Q81756	hepatitis c
4	943	99.0	3011	12	Q91RE5	Q91RE5	hepatitis c
5	942	98.8	181	12	Q91RR3	Q91RR3	hepatitis c
6	942	98.8	181	12	Q91RS1	Q91RS1	hepatitis c
7	942	98.8	181	12	Q91RQ8	Q91RQ8	hepatitis c
8	942	98.8	181	12	Q91RT1	Q91RT1	hepatitis c
9	940	98.6	181	12	Q91RR6	Q91RR6	hepatitis c
10	940	98.6	181	12	Q91RS9	Q91RS9	hepatitis c
11	939	98.5	181	12	Q91RS3	Q91RS3	hepatitis c
12	939	98.5	3011	12	Q03463	Q03463	hepatitis c
13	938	98.4	181	12	Q91RT4	Q91RT4	hepatitis c
14	938	98.4	181	12	Q91RS8	Q91RS8	hepatitis c
15	938	98.4	181	12	Q91RT3	Q91RT3	hepatitis c
16	938	98.4	181	12	Q91RS5	Q91RS5	hepatitis c

## ALIGNMENTS

## RESULT 1

ID	Q91RR8	PRELIMINARY:	PRT:	181 AA.
AC	Q91RR8:			
DC	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	NS3 protease (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.			
OX	Hepacivirus.			
OC	NCBI_TaxID=11103;			
OP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Pt.1y.			
RA	Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;			
RT	*Genetic Diversity and response to IFN of the NS3 Protease Gene from			
RT	Clinical Strains of the Hepatitis C Virus.*;			
RI	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RI	EMBL; AF369235; AAK54560.1; -			
DR	InterPro; IPR004109; HCV_NS3.			
DR	Pfam; PF02907; HCV_NS3; 1.			
KW	Protease.			

2	APITAYAAQQTGRLGCGITISLTGRDNKNGVEGEIVSTAAQTFLATCINGVCWTVYHGAG	Qy
1	APITAYAAQQTGRLGCGITISLTGRDNKNGVEGEIVSTAAQTFLATCINGVCWTVYHGAG	Db
62	TRTIASPKGPVIMQYITNVDKDILVGWPAPOGGRSLTPTCGSSSLYLVTIRHADYVPIVRRRG	Qy
61	TRTIASPKGPVIMQYITNVDKDILVGWPAPOGGRSLTPTCGSSSLYLVTIRHADYVPIVRRRG	Db

QY 122 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTMR 181  
DB 121 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180  
QY 182 S 182  
DB 181 S 181  
RESULT 2  
Q91RTS  
ID Q91RTS PRELIMINARY; PRT: 181 AA.  
AC Q91RTS  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE NS3 protease (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Pt.4;  
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayors D.L.;  
RT "Genetic Diversity and response to IPRN of the NS3 Protease Gene from  
RT Clinical Strains of the Hepatitis C Virus."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF369218; AAK34543.1; -  
DR InterPro: IPR004109; HCV\_NS3.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Protease.  
KW Protease.  
FT NON\_TER 1  
FT NON\_TER 181  
SQ SEQUENCE 181 AA; 19130 MW; 85D9186929987C35 CRC64;  
Query Match 99.2%; Score 945; DB 12: Length 181;  
Best Local Similarity 99.4%; Pred. No. 1.9e-88;  
Matches 180; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 APIATAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 61  
DB 1 APIATAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 60  
QY 62 TTTIATSPKGPVQIYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRRG 121  
DB 61 TTTIATSPKGPVQIYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRRG 120  
QY 122 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTMR 181  
DB 121 DSRGSLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180  
QY 182 S 182  
DB 181 S 181

RESULT 3  
Q81756  
ID Q81756 PRELIMINARY; PRT: 2436 AA.  
AC Q81756;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Genome polyprotein (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Choo Q.-L., Richman K., Han J.;  
RT "The nucleotide sequence of the Hepatitis C viral genome.";

RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.  
DR EMBL: M32084; AAA45677.1; -  
DR HSSP: P27958; 1A1V.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_NS5a.  
DR InterPro: IPR001650; Helicase.C.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSvir.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; helicase.C; 1.  
DR Pfam: PF00998; Viral\_RDRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS50507; RDRP\_POSITIVE; 1.  
DR PROSITE; PS50521; RDRP\_VIRAL; 1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydrolase; Nonstructural protein; Polyprotein;  
KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
FT NON\_TER 1  
FT NON\_TER 2436  
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 99.0%; Score 943; DB 12: Length 2436;  
Best Local Similarity 98.4%; Pred. No. 7.1e-87;  
Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 60  
DB 576 LAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCTVYHGA 635  
QY 61 GTRTASPKGPVQIYNTVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120  
DB 636 GTRTASPKGPVQIYNTVDQDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 695  
QY 121 GDSRGSLLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVLESLETTM 180  
DB 696 GDSRGSLLSPRPISYLKSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 755  
QY 181 RS 182  
DB 756 RS 757

RESULT 4  
Q91FE5  
ID Q91FE5 PRELIMINARY; PRT: 3011 AA.  
AC Q91FE5;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21262212; PubMed-11369872;  
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;  
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype  
RT sequence."  
RL J. Gen. Virol. 82:1291-1297(2001).







```

Db      61  TTTIAPKGPVIQMTYNTVDQDLVGVPAPOGARSITPCTCGSSDLYLVTRHADV.PVRRRG 120
QY      122  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 10
Q91RS9
ID      Q91RS9          PRELIMINARY;          PRT: 181 AA.
AC      Q91RS9;
DT      01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      NS3 protease (Fragment).
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-Pt.174;
RA      Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT      "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT      Clinical Strains of the Hepatitis C Virus.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL: AF369224; AAK54549.1; -;
DR      InterPro: IPR004109; HCV_NS3.
DR      Pfam: PF02907; HCV_NS3.1;
KW      Protease.
FT      NON_TER
FT      NON_TER
SQ      SEQUENCE 181 AA: 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match      98.6%; Score 940; DB 12; Length 181;
Best Local Similarity 98.6%; Pred. No. 6e-88;
Matches 179; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
Db      1  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60
QY      62  TTTIAPKGPVIQMTYNTVDKDLVGVPAPOGARSITPCTCGSSDLYLVTRHADV.PVRRRG 121
Db      61  TTTIAPKGPVIQMTYNTVDKDLVGVPAPOGARSITPCTCGSSDLYLVTRHADV.PVRRRG 120
QY      122  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 11
Q91RS3
ID      Q91RS3          PRELIMINARY;          PRT: 181 AA.
AC      Q91RS3;
DT      01-DEC-2001 (TReMBLrel. 19, Created)
DT      01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      NS3 protease (Fragment).
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RA      Miyakawa Y., Mayumi M.;
RA      Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RT      "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions.";
RL      J. Gen. Virol. 72:2697-2704(1991).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=93117120; PubMed=1335573;
RA      Okamoto H., Kanai N., Mishihiro S.;
RT      "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT      isolate (HC-J1) with high homology to USA isolates.";
RL      Nucleic Acids Res. 20:6410-6410(1992).
RN      [4]

```

```

RC      STRAIN-Pt.24;
RA      Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT      "Genetic Diversity and response to IFN of the NS3 protease Gene from
RT      Clinical Strains of the Hepatitis C Virus.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL: AF369230; AAK54555.1; -;
DR      InterPro: IPR004109; HCV_NS3.
DR      Pfam: PF02907; HCV_NS3.1;
KW      Protease.
FT      NON_TER
FT      NON_TER
SQ      SEQUENCE 181 AA: 19132 MW; 0BB90B5F3AB95250 CRC64;

Query Match      98.5%; Score 939; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 7.6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 61
Db      1  APITAYAOQTRGLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAG 60
QY      62  TTTIAPKGPVIQMTYNTVDKDLVGVPAPOGARSITPCTCGSSDLYLVTRHADV.PVRRRG 121
Db      61  TTTIAPKGPVIQMTYNTVDKDLVGVPAPOGARSITPCTCGSSDLYLVTRHADV.PVRRRG 120
QY      122  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
Db      121  DSRGSLSPRPISYLYKSGGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 180
QY      182  S 182
Db      181  S 181

RESULT 12
Q03463
ID      Q03463          PRELIMINARY;          PRT: 3011 AA.
AC      Q03463;
DT      01-NOV-1996 (TReMBLrel. 01, Created)
DT      01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT      01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=91013116; PubMed=2170712;
RA      Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA      Yoshizawa H.;
RT      "The 5'-terminal sequence of the hepatitis C virus genome.";
RL      Jpn. J. Exp. Med. 60:167-177(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=92044440; PubMed=1658196;
RA      Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA      Miyakawa Y., Mayumi M.;
RT      "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions.";
RL      J. Gen. Virol. 72:2697-2704(1991).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HC-J1;
RX      MEDLINE=93117120; PubMed=1335573;
RA      Okamoto H., Kanai N., Mishihiro S.;
RT      "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT      isolate (HC-J1) with high homology to USA isolates.";
RL      Nucleic Acids Res. 20:6410-6410(1992).
RN      [4]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MDLINE=941174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchauste G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200; 246-255(1994).
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D10749; BAA01582.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_NS5b.
DR InterPro; IPR002166; HCV_NS5b.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01538; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR Pfam; PF0186062; HCV_NS1; 1.
DR Pfam; PF0186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS0507; RDRP_POSITIVE; 1.
DR PROSITE; PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;
```

```
Query Match 98.5%; Score 939; DB 12; Length 3011;
Best Local Similarity 97.8%; Pred. No. 2.4e-86;
Matches 178; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPITAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCTVYHGAG 60
: |||||
DB 1026 LAPITAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCTVYHGAG 1085

QY 61 GTRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRR 120
|||||
DB 1086 GTRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180
|||||
DB 1146 GDSRGSLLSPRPISYLKSGSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1205

QY 181 RS 182
||
DB 1206 RS 1207
||
```

```
RESULT 13
Q91RT4
ID Q91RT4 PRELIMINARY; PRT; 181 AA.
AC Q91RT4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-PC.23;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369219; AAK54544.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
DR KX Protease.
DR KX NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19059 MW; 1E53C47AE8B7E5C9 CRC64;

Query Match 98.4%; Score 938; DB 12; Length 181;
Best Local Similarity 98.3%; Pred. No. 9.6e-88;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCTVYHGAG 61
|||||
DB 1 APITAYAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCTVYHGAG 60

QY 62 TRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRRG 121
|||||
DB 61 TRTIASPKGPVIOYNTVDKLVGMPAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRRG 120

QY 122 DSRGSLSPRPISYLKSGSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 181
|||||
DB 121 DSRGSLSPRPISYLKSGSGGPLLCAPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180

QY 182 S 182
||
DB 181 S 181

RESULT 14
Q91RS8
ID Q91RS8 PRELIMINARY; PRT; 181 AA.
AC Q91RS8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-Pt.176;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369225; AAK54550.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
DR KX Protease.
FT NON_TER 1 1
FT NON_TER 1
```

Job time : 36.7298 secs

```

PT NON_TER 181 181
SQ SEQUENCE 181 AA: 19114 MW: 574AC47A8ABE5D2 CRC64;

Query Match
Best Local Similarity 98.4%; Score 936; DB 12; Length 181;
Matches 178; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVS:AAOTFLATCINGVCWTYYHGAG 61
   |||
Db 1 APITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVS:AAOTFLATCINGVCWTYYHGAG 60
   |||

QY 62 TRTIASPKGPVIQMTYNVDKDLVGMWPAPOGSSITPCYCGSSDLYLVYTRHADVIIPVRRRG 121
   |||
Db 61 TRTIASPKGPVIQMTYNVDKDLVGMWPAPOGSSITPCYCGSSDLYLVYTRHADVIIPVRRRG 120
   |||

QY 122 DSRGSLSPRPISYLYKSGSGGPLLCGAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
   |||
Db 121 DSRGSLSPRPISYLYKSGSGGPLLCGAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
   |||

QY 182 S 182
Db 181 S 181

RESULT 15
Q91RT3
ID Q91RT3 PRELIMINARY; PRT; 181 AA.
AC Q91RT3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pt.11;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Meyers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF369220; AAK54545.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 181
FT NON_TER 181 181
SQ SEQUENCE 181 AA: 19116 MW: 9648807F49EB1D43 CRC64;

Query Match
Best Local Similarity 98.4%; Score 936; DB 12; Length 181;
Matches 178; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 APITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVS:AAOTFLATCINGVCWTYYHGAG 61
   |||
Db 1 APITAYAOOTRGLGCIITSLTGRDKNOVEGEVOIVS:AAOTFLATCINGVCWTYYHGAG 60
   |||

QY 62 TRTIASPKGPVIQMTYNVDKDLVGMWPAPOGSSITPCYCGSSDLYLVYTRHADVIIPVRRRG 121
   |||
Db 61 TRTIASPKGPVIQMTYNVDKDLVGMWPAPOGSSITPCYCGSSDLYLVYTRHADVIIPVRRRG 120
   |||

QY 122 DSRGSLSPRPISYLYKSGSGGPLLCGAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMR 181
   |||
Db 121 DSRGSLSPRPISYLYKSGSGGPLLCGAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMR 180
   |||

QY 182 S 182
Db 181 S 181

```

Search completed: August 30, 2003, 19:18:17

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 14.9789 Seconds  
 (without alignments)  
 1168.492 Million cell updates/sec

Title: US-09-965-594-1  
 Perfect score: 953  
 Sequence: 1 MAPITAYAOOTRGLLCIIIT.....GVAKAVDFPYESLFTTWS 162

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	943	99.0	3011	1 GNMWC3	genome polyprotein
2	939	98.5	3011	1 S40770	genome polyprotein
3	927	97.3	3011	1 GNMVCH	genome polyprotein
4	892	93.6	3010	1 GNMVTC	genome polyprotein
5	891	93.5	3010	1 GNMVTC	genome polyprotein
6	887	93.1	3010	1 GNMVCJ	genome polyprotein
7	884	92.8	3010	1 A45373	genome polyprotein
8	864	90.7	3010	1 SL8030	genome polyprotein
9	804	84.4	3034	1 JC5620	genome polyprotein
10	714	74.9	3033	1 GNMVJ8	genome polyprotein
11	712	74.7	3033	1 JQ1303	genome polyprotein
12	267.5	28.1	3005	2 T08841	polyprotein - douc
13	255.5	26.8	2970	2 T08939	polyprotein - marm
14	87	9.1	209	2 H83144	probable aromatic
15	82	8.6	452	2 I39383	angio-associated m
16	82	8.6	495	2 B71360	hypothetical prote
17	81.5	8.6	476	2 T48399	heat shock transcr
18	78.5	8.2	620	2 F83976	cytochrome-c oxida
19	78.5	8.2	981	2 T18234	beta transducin ho
20	78	8.2	239	2 H89866	serine proteinase
21	77.5	8.1	398	2 B71284	probable periplasm
22	77	8.1	904	2 A84212	hypothetical prote
23	76.5	8.0	270	2 T06118	hypothetical prote
24	76.5	8.0	868	2 H81775	aconitate hydratase
25	75	7.9	240	1 CPB0A3	procarboxypeptidas
26	74.5	7.8	415	2 S70401	zona pellucida gly
27	74.5	7.8	755	2 S23441	hypothetical prote
28	74.5	7.8	868	2 C81200	aconitate hydratase
29	74.5	7.8	911	2 JN0821	transferrin-bindin

30	74	7.8	140	2 C72705	hypothetical prote
31	74	7.8	377	2 A75335	hypothetical prote
32	74	7.8	808	2 G86208	protein F22G5.28 I
33	73.5	7.7	356	2 F90978	hypothetical prote
34	73.5	7.7	447	2 S76033	hypothetical prote
35	73.5	7.7	451	2 H82044	C4-dicarboxylate t
36	73.5	7.7	566	2 H84203	phosphate ABC tran
37	73.5	7.7	846	2 T04533	hypothetical prote
38	73.5	7.7	910	2 C81832	transferrin-bindin
39	73.5	7.7	915	2 F81196	transferrin-bindin
40	73.5	7.7	1820	2 A55494	latent transferrin
41	73	7.7	239	2 A89967	serine proteinase
42	73	7.7	354	2 T49806	hypothetical prote
43	73	7.7	433	2 H97199	htvA-like serine p
44	73	7.7	535	2 S65762	chitinase (EC 3.2.
45	72.5	7.6	249	2 A55634	granzyme M (EC 3.4

ALIGNMENTS

RESULT 1

GNMWC3

genome polyprotein - hepatitis C virus (strain HCV-1)  
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonst  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C:Accession: A39166; P00403; P00404  
 R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
 A:Title: Genetic organization and diversity of the hepatitis C virus.  
 A:Reference number: A39166; MUID:91172826; PMID:1848704  
 A:Accession: A39166  
 A:Molecule type: mRNA  
 A:Residues: 1-3011 <CHOS>  
 A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874  
 R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
 A:Reference number: P00393; MUID:92268871; PMID:1316939  
 A:Accession: P00403  
 A:Molecule type: genomic RNA  
 A:Residues: 1577-1633 <CHAS>  
 A:Cross-references: DDBJ:D10128  
 A:Experimental source: Isolates E-b16  
 A:Accession: P00404  
 A>Status: preliminary  
 A:Molecule type: genomic RNA  
 A:Residues: 1577-1633 <CH2>  
 A:Experimental source: Isolates E-b17  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 99.0% Score 943; DB 1; Length 3011;  
 Best Local Similarity 98.4% Pred. No. 5.7e-81;  
 Matches 179; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 MAPITAYAOOTRGLLCIIITSLTGDKNOVEGOIVSTAQTFLATCTINGVCWTVYHGA 60  
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATOTFLATCINGVCWTYYHGA 1085

QY 61 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120  
 |||||  
 Db 1086 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145  
 |||||

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
 |||||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182  
 ||

Db 1206 RS 1207

## RESULT 2

S40770

genome polyprotein - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: S40770; PC1285  
 R:Okamoto, H.: submitted to the EMBL Data Library, March 1992  
 A:Reference number: S40770  
 A:Accession: S40770  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <OK>  
 A:Cross-references: ENBL:D10749; NID:q221586; PIDN:BA00705.1; PID:q221587  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.: Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A:Reference number: PC1284; MUID:91013116; PMID:2170712  
 A:Accession: PC1285

A:Molecule type: genomic RNA

A:Residues: 1-513 &lt;OK&gt;

A:Cross-references: GB:D00831; NID:q221511; PIDN:BA00705.1; PID:q221512

A:Experimental source: isolate HC-J1

C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <GPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <WEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;

F:1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;

F:2014-3011/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

Query Match 98.5%; Score 939; DB 1; Length 3011;

Best Local Similarity 97.8%; Pred. No. 1.4e-80;

Matches 178; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATOTFLATCINGVCWTYYHGA 60  
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATOTFLATCINGVCWTYYHGA 1085

QY 61 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

|||

Db 1086 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180

|||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182

||

Db 1206 RS 1207

## RESULT 3

GNVWCH

genome polyprotein - hepatitis C virus (strain H)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A36814; A41546  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to GenBank, July 1992  
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus  
 A:Reference number: A36814  
 A:Accession: A36814  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <INC>  
 A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA5534.1; PID:g329738  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: com  
 A:Reference number: A41546; MUID:92052256; PMID:1658800

A:Contents: annotation

A:Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc

F:1-115/Product: capsid protein C #status predicted &lt;GPC&gt;

F:116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F:192-389/Product: major envelope protein E #status predicted &lt;WEE&gt;

F:390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F:730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F:1007-1615/Product: nonstructural protein NS3 #status predicted &lt;NS3&gt;

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: DEXH motif

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;

F:1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;

F:2014-3011/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,224

Query Match 97.3%; Score 927; DB 1; Length 3011;

Best Local Similarity 96.2%; Pred. No. 1.9e-79;

Matches 175; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATOTFLATCINGVCWTYYHGA 60  
 :|||||

Db 1026 LAPITAYAAQOTRGLGCIITSLTRDKNQVEGEVQIVSTAATOTFLATCINGVCWTYYHGA 1085

QY 61 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120

|||

Db 1086 GTRTIASPKGPVQIOMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145

QY 121 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180

|||

Db 1146 GDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

QY 181 RS 182

||

Db 1206 RS 1207

## RESULT 4

GNVWTV

genome polyprotein - hepatitis C virus (strain Taiwan)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)  
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.

virology 188, 102-113, 1992







A:Accession: J05620  
A:Molecule type: mRNA  
A:Residues: 1-3014 <CH>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
A:Experimental source: genotype 5a, which predominates in South Africa  
A:Note: The translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: major envelope protein E #status predicted <MEE>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20  
Query Match 84.4% Score 804; DB 1; Length 3014;  
Best Local Similarity 79.1% Pred. No. 9.5e-68;  
Matches 144; Conservative 22; Mismatches 16; Indels 0; Gaps 0;  
QY 1 MAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQIFLATCINGVCMVYHGA 60  
DB 1027 LAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQIFLATCINGVCMVYHGA 1086  
QY 61 GTRTASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTGSSDLVYTRHADVIPVRR 120  
DB 1087 GSKTLAGPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTGSSDLVYTRHADVIPVRR 1146  
QY 121 GDSRGLSPRPISYLVKSSGGPLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
DB 1147 GDRASLLSPRPISYLVKSSGGPLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1206  
QY 181 RS 182  
DB 1207 RS 1208  
RESULT 10  
GNVJ38  
genome polyprotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstru  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40250; P00559  
R:Okamoto, H.; Kura, K.; Okada, S.; Yanamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;  
Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep  
A:Reference number: A40250; MUID:92230232; PMID:1314459  
A:Accession: A40250  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OK>  
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609  
R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, R.; Peutherer, J.F.; Pollett, K.; Yap, P.I.  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A:Reference number: P00393; MUID:92268871; PMID:1316939  
A:Accession: P00397  
A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CH>  
A:Cross-references: DDBJ:D10134  
A:Experimental source: isolate E-b12  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: P00554; MUID:92068204; PMID:1720309

A:Accession: P00559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: major envelope protein E #status predicted <MEE>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20  
Query Match 74.9% Score 714; DB 1; Length 3033;  
Best Local Similarity 70.9% Pred. No. 3.5e-59;  
Matches 129; Conservative 26; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQIFLATCINGVCMVYHGA 60  
DB 1030 LAPITAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQIFLATCINGVCMVYHGA 1089  
QY 61 GTRTASPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTGSSDLVYTRHADVIPVRR 120  
DB 1090 GSKTLAGPKGPVQMYTNVDKLVGWPAPQGSRLTPTCTGSSDLVYTRHADVIPVRR 1149  
QY 121 GDSRGLSPRPISYLVKSSGGPLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 180  
DB 1150 DRRGALLSPRPISYLVKSSGGPLCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1209  
QY 181 RS 182  
DB 1210 RT 1211  
RESULT 11  
JQ1303  
genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonst  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C:Accession: JQ1303  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Izuka, H.; Machida, A.; Miyakaw  
J. Gen. Virol. 72, 2697-2704, 1991  
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from  
A:Reference number: JQ1303; MUID:92044440; PMID:1658196  
A:Accession: JQ1303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OK>  
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651  
A:Experimental source: isolate HC-J6 from a Japanese individual  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; t  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: major envelope protein E #status predicted <MEE>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

```
Query Match          74.7%   Score 712;   DB 1;   Length 3033;
Best Local Similarity 69.8%;   Pred. No. 5.4e-59;
Matches 127;   Conservative 29;   Mismatches 26;   Indels 0;   Gaps 0;

QY 1 MAPITAYAOQTGRLGCGITISLTGRDKNQVEGEVOIVSTAAQTFTATCINGVCWTVYHGA 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1030 LAPITAYAOQTGRLGIVVSMIGRDKTEQAGEIOVLSTVQSFGLCTTISGVLWTVYHGA 1089

QY 61 GTRTIASPKGPVQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRG 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1090 GNKTLAGSRGPVTQMTYSSAGDGLVGMPSPPGTKSLEPCTCGAVDLYLVRNADVIPARRR 1149

QY 121 GDSRGLSPRPISYLGSSGGPLCPAGHAVGIFRAAVCTRCVAKAVDFIPVESLETTM 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1150 GDRGALLSPRLTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLIDVT 1209

QY 181 RS 182
   ||
Db 1210 RS 1211

RESULT 12
T08841
polyprotein - douroucouli hepatitis GB virus A
C:Species: douroucouli hepatitis GB virus A
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T08841
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: 216486; MCID:98120818; PMID:9460920
A:Accession: T08841
A:Status: translated from GB/EMBL/DDRJ
A:Molecule type: mRNA
A:Residues: 1-3005 <ERR>
A:Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g282860C
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match          28.1%   Score 267.5;   DB 2;   Length 3005;
Best Local Similarity 33.1%;   Pred. No. 1e-16;
Matches 60;   Conservative 32;   Mismatches 78;   Indels 11;   Gaps 4;

QY 2 APITAYAOQTGRLGCGITISLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHAG 6;
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 979 APVVV-MORGGLGFTSVKTSMLGRDERRHREGSIVVLCTSTRSMGTGCVGMVMTTFHGSN 1037

QY 62 TRTIASPKGPVQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRG 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1038 ARTLAGPVGPVNCRWSPSDOVAVYPLPFGASCLPECKGCTQSWCIRN-DGALCHGRL 1095

QY 122 DSRGSLSPRPISYLGSSGGPLCPAGHAVGIFRAAVCTRCG-----AKAVDFIPVE 174
   | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1096 SKLVELDLPTFISDFRSGSSPILCDGHHVGMN-VSVLHRGKVTGTVRYVKPWETLPKD 1154

QY 175 S 175
   |
Db 1155 S 1155

RESULT 13
T08839
polyprotein - marmoset hepatitis GB virus A
C:Species: marmoset hepatitis GB virus A
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: T08839
R. Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
J. Gen. Virol. 79, 41-45, 1998
A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
A:Reference number: 216486; MUID:98120818; PMID:9460920
A:Accession: T08839
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: genomic RNA

A:Residues: 1-2970 <ERR>
A:Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match          26.8%   Score 255.5;   DB 2;   Length 2970;
Best Local Similarity 30.1%;   Pred. No. 1.4e-15;
Matches 59;   Conservative 36;   Mismatches 68;   Indels 33;   Gaps 6;

QY 2 APITAYAOQTGRLGCGITISLTGRDKNQVEGEVOIVSTAAQTFLATCINGVCWTVYHAG 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 970 APVVVH-HHGKGFVGVKTSMTGDETHGVNVVLGTSTRSMGTGCVGMVMTTFHGSN 1028

QY 62 TRTIASPKGPVQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRG 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1029 ARTLAAGPVPVNRWSSASDDVAVPLPVGAKLECKQCPQGVWVI-----RN 1077

QY 122 DSRGSLSPRPISYLGSSGGPLCPAGHAVGIFRAAVCTRCG----- 163
   | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1078 D--GALCHGTLGRTVELDLPAELCDFRSGSGSPILCDGHHVGMN-LSVLHRSRVTGIR 1134

QY 164 VAKAVDFIPVESLETT 179
   | :|||:|||||:
Db 1135 YTKPWETLPREAITHT 1150

RESULT 14
H83144
probable aromatic acid decarboxylase PA4019 [imported] - Pseudomonas aeruginosa (str.
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83144
R. Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AE004818; GB:AE004091; NID:g9950200; PIDN:AAG07406.1; GSPDB:GI
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4019
C:Superfamily: dedF protein

Query Match          9.1%   Score 87;   DB 2;   Length 209;
Best Local Similarity 26.2%;   Pred. No. 0.69;
Matches 55;   Conservative 18;   Mismatches 61;   Indels 76;   Gaps 12;

QY 8 AQTGRLGCGITISLTGRDKNQVEGEVO-IVSTAAQTFLATCINGVCWTVYHAGTRTIA 66
   || |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17 AOYGLRLDCLV-----QEEREVHFLTSKAAQLVMAT-----ETDA 53

QY 67 SPKGP-----VIQMTYNTVDKDLVGPAPQGSRLTPT-----CTCGSSDL 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 LPAKPOAMQAFLEYCGAAAGQIRVEGQD-----WMAPPASGSSAPNANAVICPCSTGTL 108

QY 106 -----YLVTRHADVIPVRRRGRDGRGSLSPR--PIS-----YLGSSGGPLCPA 148
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 SAVATGACNNLIERAADVALKER---RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA 164

QY 149 GHAVGIFRAAVCTRCVAKAVDFIPVESLET 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 --APGFYHQ---POSVEDLVDFVVARILNT 189

RESULT 15
I39383
angio-associated migratory cell protein - human
C:Species: Homo sapiens (man)
```

C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 21-Jul-2000  
C:Accession: I39383  
R:Beckner, M.E.; Krulzsch, H.C.; Stracke, M.L.; Williams, S.T.; Gallardo, J.A.; Eiotta,  
Cancer Res. 55, 2140-2149, 1995  
A:Title: Identification of a new immunoglobulin superfamily protein expressed in blood  
A:Reference number: I39383; MUID:95262124; PMID:7743515  
A:Accession: I39383  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-452 <RES>  
A:Cross-references: GB:M95627; NID:q870802; P1EN:AAA68889.1; PID:g270803  
C:Genetics:  
A:Gene: GDB:AAMP  
A:Cross-references: GDB:4573993  
A:Map position: 14q32.1-14q32.1  
C:Superfamily: unassigned WD repeat proteins; WD repeat homology  
F:148-181/Domain: WD repeat homology <WD1>  
F:414-447/Domain: WD repeat homology <WD2>

	Query Match	8.6%	Score 82;	DB 2;	Length 452;
	Best Local Similarity	25.3%	Pred. No. 5;		
Matches	42;	Conservative 13;	Mismatches 47;	Indels 64;	Gaps 9;
QY	54	MTVTHGACTRTIASPKGPVIOYTNVDKDLVGWPAPOGSRSL-----TPCTCGSSDLILY	108		
DB	197	WNEWH-----PRAPVLLAGT-ADGNTMMKVPNGDKCTFOGPNCPATCGR-----	240		
QY	109	TRHADVIPVRRR---GDSRGS-----LLSPRPISYLGSSG--GPLLCPA-----	148		
DB	241	-----VLPDGRNAVVGYPEDGIRIWDLKQGSPIHVLKTEGHQGPITCVAAQNQDGSLLIT	295		
QY	149	-----CHAVGIFR-----AAVCTRGVAKAVDFIPVESL	176		
DB	296	GSVDCQAKLVSATGKVGVGFRPETVASQPSLGEGESESNVESL	341		

Search completed: August 30, 2003, 19:20:24  
Job time : 17.9789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using iframe\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 : Search time 1764.86 seconds  
(without alignments)  
2506.388 Million ccl1 updates/sec

Title: US-09-965-594-1  
Perfect score: 953  
Sequence: 1 MAPITAYAQTRGLGCIIT.....GVAKAVDFIPVESLETIMRS 182

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO\_spool/US09965594/runat\_29082003.151919.28322/app\_query.fasta\_1.2672  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09965594 -RCGN\_1\_12630.@runat\_29082003.151919.28322 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6  
-FGAPOP=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1.\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	106	11.1	984	10	BF304699	BF304699 601888252
C 2	103.5	10.9	1199	13	BQ892487	BQ892487 AGENCOURT
C 3	101	10.6	515	14	CA023748	CA023748 HZ47E17r
C 4	101	10.6	583	12	BM374064	BM374064 EBma03_SQ
C 5	99	10.4	615	12	RJ001625	RJ001625 RJ001625
C 6	99	10.4	643	12	RJ024121	RJ024121 RJ024121
C 7	99	10.4	754	12	RJ016176	RJ016176 RJ016176
C 8	98.5	10.3	961	10	BF203316	BF203316 601865914
C 9	98.5	10.3	1031	14	CB950999	CB950999 AGENCOURT
C 10	98.5	10.3	1141	11	AK080545	AK080545 Mus muscu
C 11	97.5	10.2	779	10	BF631437	BF631437 HVSMB001
C 12	96.5	10.1	844	11	CNS09045	AK080545 Single re
C 13	96	10.1	701	10	BF863244	BF863244 963042C02
C 14	96	10.1	846	10	BF182274	BF182274 601804028
C 15	95.5	10.0	901	10	BF307233	BF307233 601891502
C 16	95.5	10.0	958	10	BG420860	BG420860 602452062
C 17	95	10.0	407	9	AW785806	AW785806 117260 MA
C 18	95	10.0	460	14	CB883286	CB883286 HQ01M02w
C 19	94.5	9.9	539	28	BH349665	BH349665 CH230-65E
C 20	94	9.9	582	14	CB286751	CB286751 CMD45_C08
C 21	94	9.9	1052	10	BG398041	BG398041 602439571
C 22	94	9.9	1283	13	BQ709745	BQ709745 AGENCOURT
C 23	93.5	9.8	736	12	BI768830	BI768830 603057734
C 24	93.5	9.8	938	13	BQ894657	BQ894657 AGENCOURT
C 25	93.5	9.8	993	9	AL555424	AL555424 AL555424
C 26	93	9.8	457	29	CNS02XOL	AL219930 Tetraodon
C 27	93	9.8	470	13	BQ758584	BQ758584 EBma07_SQ
C 28	93	9.8	1018	12	BQ054587	BQ054587 AGENCOURT
C 29	93	9.8	1291	10	BE622016	BE622016 601440668
C 30	92.5	9.7	655	14	CB868789	CB868789 HC09614w
C 31	92.5	9.7	1008	12	BI755608	BI755608 603027112
C 32	92.5	9.7	1411	11	BC020343	BC020343 Homo sapi
C 33	92	9.7	756	14	CD348815	CD348815 UI-M-FY0-
C 34	92	9.7	832	10	BG387051	BG387051 602454749
C 35	92	9.7	871	10	BG178418	BG178418 602330206
C 36	92	9.7	898	29	CNS01VR5	AL169466 Tetraodon
C 37	92	9.7	963	10	BF704182	BF704182 602355566
C 38	92	9.7	1329	13	BQ960995	BQ960995 AGENCOURT
C 39	92	9.7	1640	10	BF180599	BF180599 601808704
C 40	91.5	9.6	422	14	CB763743	CB763743 AMGNNUC:S
C 41	91.5	9.6	539	10	BE757615	BE757615 212104 MA
C 42	91.5	9.6	641	9	AU127824	AU127824 AU127824
C 43	91.5	9.6	691	10	BB632604	BB632604 BB632604
C 44	91.5	9.6	701	14	CD262790	CD262790 p5MA019XE
C 45	91.5	9.6	844	12	BI198486	BI198486 602760491

ALIGNMENTS

RESULT 1  
BF304699/c  
LOCUS  
DEFINITION BF304699 984 bp mRNA linear EST 21-NOV-2000  
601888252F1 NIH\_MGC\_17 Homo sapiens CDNA clone IMAGE:412276 5',  
mRNA sequence.  
ACCESSION BF304699  
VERSION BF304699.1 GI:11251586  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 984)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: A7CC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LICM1005 row: g column: 13  
 High quality sequence stop: 646.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..984  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:412276"  
 /tissue\_type="rhodomyosarcoma"  
 /lab\_host="NIH-MGC-17"  
 /clone\_lib="NIH-MGC-17"  
 /note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 133 a 329 c 351 g 171 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.34 Length: 984  
 Score: 105.00 Matches: 33  
 Percent Similarity: 45.24% Conservative: 5  
 Best Local Similarity: 39.29% Mismatches: 24  
 Query Match: 11.12% Indels: 22  
 DB: 10 Gaps: 5  
 US-09-965-594-1 (1-182) x BF304699 (1-984)  
 QY 86 TrpProAlaIrpGInGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 104  
 Db 646 TGGCCAGTCACCGCATTCCTCCGAGAGAGACCGTGTACCTGC----- 599  
 QY 105 LeuTyrLeuValThrArgHisAlaAspValIleProValAlaArgArgGlyAspSerArg 124  
 Db 598 -----ACCAGSCAGCAGCGAACAATACATGCAAGGACCTGCT---TCCCGC 554  
 QY 125 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySer 139  
 Db 553 GGGCCCTCTTGTGGGGAAGACCTCGATGTGTCCAAAGCTCCGCTGCTCTACTGGAAGT 494  
 QY 140 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 159  
 Db 493 CGGCAGCTCCGGTCAGGTGCAGC-----TTCAGCGCCCGGGG 455  
 QY 160 CysThrArgGly 163  
 Db 454 TCCGCCCGAGGA 443  
 RESULT 2  
 BQ892487  
 LOCUS BQ892487 1199 bp mRNA linear EST 16-AUG-2002  
 DEFINITION AGENCOURT\_8417538 Lupski-sympathetic\_trunk Homo sapiens cDNA clone  
 IMAGE:6192708 5', mRNA sequence.  
 ACCESSION BQ892487  
 VERSION BQ892487.1 GI:22284501  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1199)  
 NIH-MGC http://mgi.nci.nih.gov/  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone Distribution: Agencourt Bioscience Corporation  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM3595 row: c column: 13  
 High quality sequence start: 57  
 High quality sequence stop: 394.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..1199  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6192708"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski-sympathetic-trunk"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
 NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCACCGCTCCG-3' and  
 5'-GACTAGTCTAGTCGCGCGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine); available through Life  
 Technologies."  
 BASE COUNT 255 a 362 c 343 g 211 t 28 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.79 Length: 1199  
 Score: 103.50 Matches: 41  
 Percent Similarity: 37.42% Conservative: 17  
 Best Local Similarity: 26.45% Mismatches: 53  
 Query Match: 10.86% Indels: 44  
 DB: 13 Gaps: 6  
 US-09-965-594-1 (1-182) x BQ892487 (1-1199)  
 QY 54 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle 73  
 Db 484 TGGCATCATTATTAATAAAGGTGCTCTGTTAATCATGCGCCACCGCCCGCTGATA 543  
 QY 74 GlnMetTyrTrpAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArg 93  
 Db 544 CTTCATTACCATGTGACAGTGACTTT----- 573  
 QY 94 SerLeuThrProCysThr-----CysGlySerSerAsp 104  
 Db 574 ---TGTGCTGCTGCACAGCACCCCATGACCGATGTGGCTTATGTGGAACGCGAG 630  
 QY 105 LeuTyr-LeuValThr-----ArgHisAlaAspValIleProValArg----- 118  
 Db 631 CGGTCATTGGCCCATCCCTCTCTATAAAAACAGCCCAAGCTGCTTCATCGGCGGCT 690  
 QY 119 -----ArgArgGlyAspSerArgGlySerLeuLeu-- 128  
 Db 691 GGGGTGTTGGCAGCGCGNAGCGGGTGGGCGATGGTAGGACTCGGGGGCGCGATTCTGT 750  
 QY 129 -----SerProArgProIleSerTyrLeuLys-----GlySerSerG1 141



Non-normalised library, directionally cloned into pSPORC-  
Derived from maternal tissue dissected from developing  
grains (8 days post anthesis) in glasshouse grown barley  
plants. Developed as part of the barley transcriptome  
resources of HBSRC/SEIRAD funded cereal IGF (investigating  
Gene Function) project."

BASE COUNT 92 a 176 c 174 g 141 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.6 Length: 583  
Score: 101.00 Matches: 44  
Percent Similarity: 41.96% Conservative: 16  
Best Local Similarity: 30.77% Mismatches: 53  
Query Match: 10.60% Indels: 31  
DB: 12 Gaps: 6

US-09-965-594-1 (1-182) x BM374064 (1-583)

QY 8 AlaGlnGlnThrArgGlyLeuLeuGlyCysIleThrSerLeuThrGlyArgAspLys 27  
||||| ||||||| ||||| ||||| |||||  
DB 11 GCACAGCGCGCGCGGCTAGCTGTATTCACGTGTCACACACACGACCAT 70  
QY 28 AsnGlnValGlu-----GlyGluValGlnIleValSer 38  
||||| ||||||| |||||||  
DB 71 CATCAATACC-TCCGCGAGAGTTTCAGCGATATTTCTGGAGAGCTTGGCGGGAAGA 129  
QY 39 ThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrVal----- 56  
||||| ||||| ||||| ||||| |||||  
DB 130 GGAGCGCG-----GCATGGTGTCTTCTGCTTCTGGTGAGACGCGGCGGA 177  
QY 57 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetLys 76  
||||| ||||||| ||||||| |||||||  
DB 178 AGTGTGGAGCAGGAGCGCGCGCATCTCTCGCGGTGCGCGGCTGCGCGCAGCG 237  
QY 77 ThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThr 96  
||||| ||||| ||||| ||||| |||||  
DB 238 TGGCGGACA-----TGGAGCGCGCACCGGCTCTGTACTCTCTCTCA 279  
QY 97 ProCysThrCys-----GlySerSerAsp-IleuThrLeuValThrArgHisAlaAspVa 114  
||||| ||||| ||||| ||||| |||||  
DB 280 CGTGCACCGCGTCACTCGCGCGGCGCATCTGCACCTTCTGCGCGCATGCTCAAGT 339  
QY 114 IleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSe 134  
||||| ||||||| ||||||| |||||||  
DB 340 CTTACCG-----CCACTACCGGCTCTACTAGACCGCGGTGCTGAC 381  
QY 134 rTyrLeu 136  
DB 382 GCTGATT 388

RESULT 5  
BJ001625/c 615 bp mRNA linear EST 05-DEC-2001  
LOCUS  
DEFINITION BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',  
mRNA sequence.

ACCESSION BJ001625  
VERSION BJ001625.1 GI:17364516  
KEYWORDS EST.  
SOURCE Oryzias latipes (Japanese medaka)  
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 615)  
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.  
TITLE Medaka Est Project in Takeda's lab  
JOURNAL Unpublished  
COMMENT Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsunien@nig.ac.jp.

FEATURES  
Location/Qualifiers  
Source

1..615  
/organism="Oryzias latipes"  
/mol\_type="mRNA"  
/strain="Hd-rP"  
/db\_xref="taxon:8090"  
/clone="MF01SSA025C02"  
/sex="mixture of female and male"  
/tissue\_type="whole embryo"  
/dev\_stage="segmentation stage 20 - 25"  
/clone\_lib="MF01SSA cDNA"

BASE COUNT 140 a 166 c 165 g 144 t  
ORIGIN

Alignment Scores:

Pred. No.: 11 Length: 615  
Score: 99.00 Matches: 42  
Percent Similarity: 33.77% Conservative: 9  
Best Local Similarity: 27.81% Mismatches: 50  
Query Match: 10.39% Indels: 50  
DB: 12 Gaps: 7

US-09-965-594-1 (1-182) x BJ001625 (1-615)

QY 27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46  
||||| ||||||| ||||| ||||| |||||  
DB 511 AAAATAGCTAGTAAACCAAGACACAGATCCACACACATGTTCTGGTCTTACGGGCT 452  
QY 47 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 66  
||||| ||||||| ||||||| |||||||  
DB 451 -----TGTGGAGAACCTATCACAGTTCTCTCTTAGAGCAACGGCA 410  
QY 67 SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 81  
||||| ||||||| |||||||  
DB 409 GCTCTGGCGGCGGAGAGCTCTTGGCGCAGTTGTG----- 374  
QY 82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 97  
||||| ||||| ||||| |||||  
DB 373 -----ACTCGTGAGGAGAGAGCGTCAACCGGAGGTAGG 335  
QY 98 -----CysThrCysGlySerSerAspLeuTyrLeuValThrArg----- 110  
||||| ||||||| ||||||| |||||||  
DB 334 CTCCAGGATGGGATGTGGCTCTGCT-----TTGGTCTCTCTCTCTCTCTCTCTCT 284  
QY 111 -----HisAlaAspValIleProValArgArgArgGlyAspSer 123  
||||| ||||| ||||| |||||  
DB 283 TCTTCTCACTGACCTTCCACATCCAGTGTGCGGCGGCTGTCTGACGGGTGATGGG 224  
QY 124 ArgGlySerLeuLeuSerProArg-----ProlleSerTyrLeuLysGlySer 140  
||||| ||||| ||||| |||||  
DB 223 AGAGCGCGGACAGCGAGTGGGGGTGAATCTCTGACGAGCGTCTTACGGCGGATCA 164  
QY 141 GlyGlyProLeuLeuCysProAlaGlyHisAla 151  
||||| ||||||| ||||||| |||||  
DB 163 GGAGGACGACTCGCTGCAGAGCCCTCTCTCTGCA 131

RESULT 6  
BJ024121  
LOCUS  
DEFINITION BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',  
mRNA sequence.  
ACCESSION BJ024121  
VERSION BJ024121.1 GI:17377389  
KEYWORDS EST.  
SOURCE Oryzias latipes (Japanese medaka)  
ORGANISM

Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

```

REFERENCE 1 (bases 1 to 643)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES             source
    source
1..643
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA143D12"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 11.7 Length: 643
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 10.39% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-1 (1-182) x BJ024121 (1-643)
Qy 27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
||||| ||||| ||||| ||||| |||||
Db 242 AAAAATGACGTAGAACCAACAAACACACATCCACACACATGTTCTGTCTACGGGCT 301
Qy 47 ThrCysileAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrileAla 66
||||| ||||| ||||| ||||| |||||
Db 302 -----TGTGGAGACCTATCACAGTTCTCTGCTTTAGACCAACGGCA 343
Qy 67 SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 81
||| ||| ||||| ||||| |||||
Db 344 GCTCTGGGGGGGAGGAGCTCTGGCCAGTTGTG----- 379
Qy 82 AspLeuValClyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 97
||| ||| ||||| ||||| |||||
Db 380 -----ACTCTGGAGACGAGAGCGCTCACCCGGAGCTGTAGG 418
Qy 98 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 110
||| ||| ||||| ||||| |||||
Db 419 CTGACGGGATCGGATGCTGCTGCT-----TTGGTTCTCTCTCTCTCGATCA 469
Qy 111 -----HisAlaAspValIleProValArgArgGlyAspSer 123
||| ||| ||||| ||||| |||||
Db 470 TCTTCTCACTGACCTTCCATCCAGGTGTGCCAGCGCTGGTCTGACGGGTGATGGG 529
Qy 124 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySer 140
||||| ||| ||||| ||||| |||||
Db 530 AGAGCGCGGACAGCAGCTCGGGGTGAATCTCTGCAGGACGCTTTCACGCGGATCA 589
Qy 141 GlyCysProLeuLeuCysProAlaGlyHisAla 151
||||| ||| ||||| ||||| |||||
Db 590 GGAGGACCGACTCGCTGCAGAGCTCTGCTGCA 622

RESULT 7
LOCUS BJ016176 754 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
mRNA sequence.
ACCESSION BJ016176

```

```

VERSION BJ016176.1 GI:17376695
KEYWORDS RST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
REFERENCE 1 (bases 1 to 754)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES             source
    source
1..754
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-rR"
/db_xref="taxon:8090"
/clone="MF01SSA025C02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

BASE COUNT 194 a 181 c 181 g 198 t
ORIGIN

Alignment Scores:
Pred. No.: 14.4 Length: 754
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 10.39% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-1 (1-182) x BJ016176 (1-754)
Qy 27 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 46
||||| ||||| ||||| ||||| |||||
Db 242 AAAAATGACGTAGAACCAACAAACACACATCCACACACATGTTCTGTCTACGGGCT 301
Qy 47 ThrCysileAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrileAla 66
||||| ||||| ||||| ||||| |||||
Db 302 -----TGTGGAGACCTATCACAGTTCTCTGCTTTAGACCAACGGCA 343
Qy 67 SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 81
||| ||| ||||| ||||| |||||
Db 344 GCTCTGGGGGGGAGGAGCTCTGGCCAGTTGTG----- 379
Qy 82 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 97
||| ||| ||||| ||||| |||||
Db 380 -----ACTCTGGAGGACGAGAGCGCTCACCCGGAGCTGTAGG 418
Qy 98 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 110
||| ||| ||||| ||||| |||||
Db 419 CTGACGGGATCGGATGCTGCTGCT-----TTGGTTCTCTCTCTCTCGATCA 469
Qy 111 -----HisAlaAspValIleProValArgArgGlyAspSer 123
||| ||| ||||| ||||| |||||
Db 470 TCTTCTCACTGACCTTCCATCCAGGTGTGCCAGCGCTGGTCTGACGGGTGATGGG 529
Qy 124 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySer 140
||||| ||| ||||| ||||| |||||
Db 530 AGAGCGCGGACAGCAGCTCGGGGTGAATCTCTGCAGGACGCTTTCACGCGGATCA 589
Qy 141 GlyGlyProLeuLeuCysProAlaGlyHisAla 151
||||| ||| ||||| ||||| |||||

```





```

Db      452 TTG---GGCACACTGTGGTCCGTGGCACAT-----ATCGATCGCCCT 490
QY      69  LysGlyProValIleGlnMetYrThrAsnValAspLysAspLeuValClyTrpProAla 88
Db      491 AAGAGGCCCTTTACAAAACACT-----AACCTCCTCGCTGGCGCTGCC 535
QY      89  Pro-----GlnGlySerArgSerLeuThrProCysThrCysGly 101
Db      536 AT-GTTGGCAAAAGAACCGTTTGTGGCTTCGGCCCTCTCGCCGCCCAATTGGGA 594
QY      102 SerSerAspLeuYrLeuValThrArgHisAlaAsp-VaiIleProValArgArgGly 121
Db      595 ACCAGTGGC-----ACCACCATGGGGTGTGTGTCCCGCGCTCTCCCGTGG 642
QY      121 YAspSerArgGlySerLeuLeuSerProArgProIleSerYrLeuLysGlySerSerGly 141
Db      643 GCAATTACAAACNCCCTTAACCGTCCCTCCCAACATATTCTTCAAGCGTCTCTGA 702
QY      141 y-----GlyProLeuLeuCysProAlaGlyHisAlaValGly 153
Db      703 TTTCCTTAAGTCCCGCCCTTTGTGTACCCACGACCATTTGTGGGA 748

RESULT 10
AK080545
LOCUS
DEFINITION
Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:A730082L10 product:weakly similar to zinc
finger protein (fragment) [Mus musculus], full insert sequence.
ACCESSION
AK080545
VERSION
1 GI:26348600
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
2049374
MEDLINE
11042159
PUBMED
10349636
REFERENCE
2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
2049374
MEDLINE
11042159
PUBMED
10349636
REFERENCE
3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Toqawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,I., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (KISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
20530913
MEDLINE
11076861
PUBMED
11076861
REFERENCE
4
Kawai,J., Shingawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,T.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,

```

```

Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Machionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Suzuki,S.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohlsuki,S.,
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
REFERENCE
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase 1 & II team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1141)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
location/Qualifiers
1. 1141
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:A730082L10"
/db_xref="taxon,10090"
/clone="A730082L10"
/tissue_type="cerebellum"
/collection="RIKEN full-length enriched mouse cDNA library"
/dev_stage="7 days neonate"
<1. 587
/note="unnamed protein product: putative
weakly similar to zinc finger protein (fragment) [Mus
musculus] (PIR)I48722, evidence: FASTY, 50.7%ID,
57.6%length, match=601"
/codon_start=3
/db_xref="GI:26348601"
/translation="DSCLPAAAPGSRLLTPRGDGFLEKLSAAARVAGPGSVAFGVS
TVRGAAQAFQGRVACRRSEGLCSRRPRRQHPVPVPHVYGLSGRRIPPPAGE
AQAAGAQOQPHPPGPRPGTGVVPPQGAAGLLPALAARQVPPVGRGPRAPRHS
PKPVPALGFGSGGPGAPPAPLLAPANGRSVGLAL"
1118. 1123
/note="putative"
polya_signal

```



623	ATCTTTACCGGTCGGTGCCACCTCTCTTACAGCTGGCTGGATCGCATCAGAACGATA	564
66	AlaSerProLys	77
563	CCCGGTTCAAGAGCTGCTGATGCTGCTGGCGAGCGGGCCCTCTTTGCGC	513
78	AsnValAspLysAspLeuValGlyTTPProAlaProGlnGlySerArgSerLeuThrPro	97
512	TGTCGCTGACTTGTGTCTCCGCTCGAA	498
98	CysThrCys	113
497	TGTCGCTGACTTGTGTCTCCGCTCGAA	456
114	ValleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle	133
435	ACGAGCGCTGGCAATCTGTTCACCCAGTATCGTA	420
134	SerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGly	153
419	GCAATTCG	363
154	IlePheArgAlaAlaValCysThrArgGlyValAlaLys	169
362	ACATTCCGCCGA	309
170	PheIleProVal	173
308	GGGTTCCTGTC	297
RESULT 13	BF863244	
LOCUS	963042C02.xl C. reinhardtii CC-1690, Stress Condition I, normalized	
DEFINITION	Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.	
ACCESSION	BF863244.1 GI:12253388	
VERSION	EST.	
KEYWORDS	Chlamydomonas reinhardtii	
SOURCE	Chlamydomonas reinhardtii	
ORGANISM	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;	
REFERENCE	1 (bases 1 to 701)	
AUTHORS	Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C., Lefebvre, P., Mcdermott, J.P., Shrager, J., Silflow, C. and Stern, D.	
TITLE	Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 3	
JOURNAL	Unpublished	
COMMENT	Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu.	
FEATURES	Location/Qualifiers	
source	1..701	
	/organism="Chlamydomonas reinhardtii"	
	/mol_type="mRNA"	
	/strain="CC-1690 wild type mt+ 2lgr"	
	/db_xref="taxon:3055"	
	/clone_lib="C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"	
	/note="Vector: pluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites.	

phuescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with EXAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t  
ORIGIN

Alignment Scores:  
Pred. No.: 25.4 Length: 701  
Score: 96.00 Matches: 32  
Percent Similarity: 40.71% Conservative: 14  
Best Local Similarity: 28.32% Mismatches: 45  
Query Match: 10.07% Indels: 22  
DB: 10 Gaps: 4

US-09-965-594-1 (1-182) x BF863244 (1-701)

Qy 57 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 72  
Db 171 CACCACCATACCTGTCTCTCAGGTGCTCACACCAAAATATGCCATACGGCCACTA 230  
Qy 73 IleGlnMetThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 92  
Db 231 ACAAGTTACTATACCG-----AAGGACACCGCGCTTGCCACCCCTTGAGCGG 284  
Qy 93 ArgSerLeuThrProCysThrCysGlySerSerAspLeuValThrArgHisAla 112  
Db 285 AGAAGCCGACCGTGTCTCTGGGTCTATCGCATCTGCTCAATCTCCGCTATCAG 344  
Qy 113 AspValIle-----ProValArgArgArgGlyAspSerArg----- 124  
Db 345 GAGATCATTTGGCATGTGCTTTAGTCACCCCAAGAGAGCGCTGGGAGTGGGATTATAA 404  
Qy 125 -----GlySerLeuSerProArgProArgProLysSerTyrLeu 136  
Db 405 GAAGGCGACGGGAATTCGGTTTCGGAAGATGAGCGCCCAAGTCTGACCAAGTGCTA 464  
Qy 137 LysGlySerGlyGlyProLeuLeuCysProAlaGly 149  
Db 465 CTCACGACGACAAATGGAGCCCTTCGGGTGTGGCGGT 503

RESULT 14  
BF182274/c 846 bp mRNA linear EST 31-OCT-2000  
LOCUS 601804028F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',  
DEFINITION mRNA sequence.

ACCESSION BF182274.1 GI:11050416

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM9308 row: g column: 07

High quality sequence stop: 696.

Location/Qualifiers

1..846

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:4035102"

/tissue\_type="tumor, gross tissue"

/dev\_stage="7 months"

/lab\_host="DH10B"

/clone\_lib="NCI\_CGAP\_Mam5"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 176 a 218 c 241 g 210 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 32.5 Length: 846

Score: 96.00 Matches: 46

Percent Similarity: 47.20% Conservative: 13

Best Local Similarity: 36.80% Mismatches: 36

Query Match: 10.07% Indels: 31

DB: 10 Gaps: 9

US-09-965-594-1 (1-182) x BF182274 (1-846)

Qy 59 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAs 78

Db 757 GGTTCCTCTACCAAGAACGCTGGATGAAGAAAGACCA-----CATCCTTC 710

Qy 78 nValAspLysAspLeuValGlyTrp-----ProAlaProGln 90

Db 709 GGTTCTTCTACCAAGTGGGCTGGAGAAAGTGAACACAGACAGACGCTCCCTCA 650

Qy 90 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVa 108

Db 649 GTCCCCACGCTCTAGTAGTCTAGACAAAGTGTCTGCTGGA----- 610

Qy 108 lThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLe 128

Db 609 -ACTAGACACACCT--GTAATCCAGGAGAACGCTGGAGAACACAGAGGACTCC---CT 556

Qy 128 userProArgProLysSerTyrLeuLysGlySerSerGlyGlyProLeu---LeuCysPr 147

Db 555 GACCCACCTCCC---TCCGTCTCTAGCGGACCTCTCTCGGCCCCACCTCCCTCTGTCC 499

Qy 147 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 162

Db 498 TAGTGGCACCCTCTCCCGACGACAGCAGACTGTACTCCCTTTGGCCCTCTGCACTCT 439

Qy 162 gGlyValAlaLys 166

Db 438 TGGGATGACTGAG 426

RESULT 15

BF307233

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

BF307233

601891502F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4137145 5',

mRNA sequence.

BF307233.1 GI:11254342

EST.

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LUCM1044 row: c column: 02  
 High quality sequence start: 6  
 High quality sequence stop: 684  
 Location/Qualifiers  
 1. .901  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4137145"  
 /tissue\_type="rhabdomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_17"  
 /note="Organ: muscle; Vector: pOH7; Site:1: PcoRI;  
 Site:2: XhoI; cDNA made by ol:go-dt priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 144 a 267 c 329 g 161 t

ORIGIN

#### Alignment Scores:

Pred. No.:	39.5	Length:	901
Score:	95.50	Matches:	37
Percent Similarity:	38.46%	Conservative:	8
Best Local Similarity:	31.62%	Mismatches:	28
Query Match:	10.02%	Indels:	45
DB:	10	Caps:	5

US-09-965-594-1 (1-182) x BF307233 (1-901)

Qy	52	ValCysTrpThr-Va	LyHis-----GlyAlaGlyThrArgThrIleAl	66
Db	620	ATGTGCTGGAGCGGTCCCGCACGCCCATCTCTGAGCGGGGTCCCGCACACACACAGCGTGG	679	
Qy	66	aSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTr	86	
Db	680	TGGCAGCAGGTGGAGTGTGCA-----GTGTCAGGATG	715	
Qy	86	pProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy	106	
Db	716	GCCTGGCCCATCCGG-----	731	
Qy	106	rLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlyse	126	
Db	732	-----GTGACCTTCGTCTCAGGGCGCTTGGCGGGGCTTC	766	
Qy	126	rLeuLeuSerProArgProIleSerTyrLeuLysGlySerScrGlyGlyProLeuLeuCy	146	
Db	767	CCITTTGGGTCCTGA-----CGGGCATCTCTCCAGGGCGCGTGGACTG	810	
Qy	146	sProAlaGlyHisAlaValGly-----IlePheArgAlaAlaValCys	160	
Db	811	TCCGGCGGTTCGGCCCGGGCGGCACACAGGCTCCGGCGGGGCTGTGC	859	

Search completed: August 31, 2003, 04:27:18  
 Job time : 1773.86 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 : Search time 44.1697 seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: us-09-965-594-12  
Perfect score: 1021  
Sequence: 1 MKKGSVIVGRIVLNGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1021	100.0	195	21	AA15220
2	998	97.7	197	21	AA15221
3	982	96.2	195	21	AA15212
4	981	96.1	197	21	AA15222
5	959	93.9	197	21	AA15226
6	951	93.1	197	21	AA15223
7	939	92.0	197	21	AA15224
8	929	91.0	197	21	AA15225
9	899.5	88.1	191	21	AA14728

10	898.5	88.0	2816	14	AA134009
11	895.5	87.7	3011	14	AA140120
12	894.5	87.6	1766	10	AA192041
13	894.5	87.6	1786	10	AA190158
14	894.5	87.6	2261	10	AA190164
15	894.5	87.6	2301	10	AA192047
16	894.5	87.6	2436	10	AA192050
17	894.5	87.6	2436	10	AA192088
18	894.5	87.6	2772	21	AA18540
19	894.5	87.6	2894	16	AA170230
20	894.5	87.6	2955	20	AA114975
21	894.5	87.6	2955	21	AA18541
22	894.5	87.6	3011	13	AA121519
23	894.5	87.6	3011	14	AA131621
24	894.5	87.6	3011	17	AA190931
25	894.5	87.6	3011	18	AA134480
26	894.5	87.6	3011	19	AA140038
27	894.5	87.6	3011	23	AA122049
28	894.5	87.6	3011	23	AA184597
29	894	87.6	182	21	AA15211
30	894	87.6	609	15	AA151170
31	894	87.6	631	18	AA131884
32	894	87.6	686	23	AA18689
33	894	87.6	686	23	AA178377
34	894	87.6	686	24	ABG72261
35	893.5	87.5	3011	15	AA166995
36	892	87.4	631	20	AA193482
37	891.5	87.3	665	20	AA124943
38	891.5	87.3	2435	13	AA125135
39	891.5	87.3	2436	13	AA128582
40	890.5	87.2	2772	11	AA108123
41	890	87.2	532	23	AA121847
42	890	87.2	632	23	AA119905
43	890	87.2	686	23	AA121837
44	890	87.2	686	23	AA121838
45	890	87.2	686	23	AA121839

ALIGNMENTS

RESULT 1  
AA15220  
ID AA15220 standard; protein; 195 AA.  
AC AA15220;  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutain.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN W0200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI; 2000-465976/40.  
DR N-PSDB; AAA73329.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

HCV-1 polyprotein.  
HCV genomic amino  
Sequence encoded i  
Protein sequence o  
Peptide encoded by  
Sequence encoded i  
Sequence encoded i  
Peptide encoded by  
Protein encoded by  
Composite hepatiti  
Amino acid sequenc  
Polyprotein encode  
Compiled HCV sequ  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
HCV polyprotein.  
HCV polyprotein.  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
HCV-1 NS3/4a mutan  
Hepatitis C virus  
HCV-1 NS3/4a confo  
Hepatitis C virus  
HCV NS3 protein.  
HCV polyprotein  
HCV polypeptide 1.  
HCV amino acid seq  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus  
Hepatitis C virus

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX  
 PS Claim 23; Fig 12; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 195 AA;

Query Match 100.0%; Score 1021; DB 21; Length 195;  
 Best Local Similarity 100.0%; Pred. NO. 2.7e-98;  
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKKGSVIIVGRIVLNGAYAAQTRGEGGCOETSGTRGKNOVEGEVQIVSTAAQTFLATC 60  
 DB 1 MKKKGSVIIVGRIVLNGAYAAQTRGEGGCOETSGTRGKNOVEGEVQIVSTAAQTFLATC 60  
 QY 61 INGVCWTVYHGAGTRTIIASPKGPVIOYTNVDKDLVGPAPQGSRSLSLTCTCGSSDLY 120  
 DB 61 INGVCWTVYHGAGTRTIIASPKGPVIOYTNVDKDLVGPAPQGSRSLSLTCTCGSSDLY 120  
 QY 121 TRHADVIPVRRGDSRGLSPRPISYLGKSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 DB 121 TRHADVIPVRRGDSRGLSPRPISYLGKSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 QY 181 DFIPVESLETTMRSP 195  
 DB 181 DFIPVESLETTMRSP 195

## RESULT 2

AAB15221  
 ID AAB15221 standard; protein: 197 AA.

XX  
 AC AAB15221;

XX  
 DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #3.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX liver failure; liver cancer; mutant; mutain.

XX Hepatitis C virus.

OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

DR N-PSDB; AAA73330.

XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX  
 PS Claim 23; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 97.7%; Score 998; DB 21; Length 197;  
 Best Local Similarity 98.0%; Pred. NO. 7e-96;  
 Matches 193; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 1 MKKKGSVIIVGRIVLNG--AYAQOTRGECCOETSGTRGKNOVEGEVQIVSTAAQTFLA 58  
 DB 1 MKKKGSVIIVGRIVLNGSDTAYAAQTRGECCOETSGTRGKNOVEGEVQIVSTAAQTFLA 60  
 QY 59 TCINGVCWTVYHGAGTRTIIASPKGPVIOYTNVDKDLVGPAPQGSRSLSLTCTCGSSDLY 118  
 DB 61 TCINGVCWTVYHGAGTRTIIASPKGPVIOYTNVDKDLVGPAPQGSRSLSLTCTCGSSDLY 120  
 QY 119 LVTRHADVIPVRRGDSRGLSPRPISYLGKSGGGLLCPAGHAGVIFRAAVCTRGVAK 178  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGKSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 QY 179 AVDFIPVESLETTMRSP 195  
 DB 181 AVDFIPVESLETTMRSP 197

## RESULT 3

AAB15212

ID AAB15212 standard; protein: 195 AA.

XX  
 AC AAB15212;

XX  
 DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #1.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX liver failure; liver cancer.

XX Hepatitis C virus.

OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

DR N-PSDB; AAA73328.

DR



```

XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Example 2; Fig 10; 66pp; English.
XX
CC The present sequence is a fusion protein created using the Hepatitis C
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both
CC essential for the replication of the virus, acting to cleave its
CC replicative proteins from the polyprotein produced from the HCV genome.
CC Inhibitors of the two proteins should be effective as antiviral
CC treatments of HCV infection. This is useful as HCV can lead to chronic
CC liver disease such as cirrhosis, liver failure and liver cancer. The
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion
CC proteins which can be used to identify inhibitors of this type, as well
CC as enabling structural studies of the protease and protease:inhibitor
CC complexes.
XX
SQ Sequence 195 AA;
Query Match 96.2%; Score 982; DB 21; Length 195;
Best Local Similarity 97.4%; Pred. No. 3.2e-94;
Matches 190; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MKKKGSVVIVGRVILNCAVAQOTRGEGCQETSGTGRDKNQVGEVQIVSTAAQTFLATC 60
DB 1 MKKKGSVVIVGRVILNCAVAQOTRGLIGCIITSLTGRDKNQVGEVQIVSTAAQTFLATC 60
QY 61 INGVCMVTYVHGAGTRTIAAPKGPVIOYMTNVNDKDLVGPAPQGSRLTPTCTCGSSDLY 120
DB 61 INGVCMVTYVHGAGTRTIAAPKGPVIOYMTNVNDKDLVGPAPQGSRLTPTCTCGSSDLY 120
QY 121 TRHADVIPVRRRGRSGSLSPRPISYLKSGSGPLLCPCPAGHVGIFRAAVCTRGVAK 180
DB 121 TRHADVIPVRRRGRSGSLSPRPISYLKSGSGPLLCPCPAGHVGIFRAAVCTRGVAK 180
QY 181 DFIPVESLETTMRSP 195
DB 181 DFIPVESLETTMRSP 195
RESULT 4
AAB15222
ID AAB15222 standard; protein: 197 AA.
XX
AC AAB15222;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #4.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; muteln.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
N-PSDB; AAA73331.

```

```

XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 23; Fig 14; 66pp; English.
XX
CC The present sequence is a mutated version of a fusion protein created
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These
CC proteins are both essential for the replication of the virus, acting to
CC cleave its replicative proteins from the polyprotein produced from the
CC HCV genome. Inhibitors of the two proteins should be effective as
CC antiviral treatments of HCV infection. This is useful as HCV can lead to
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.
CC The present invention concerns a number of NS3 mutants and NS3-NS4A
CC fusion proteins which can be used to identify inhibitors of this type, as
CC well as enabling structural studies of the protease and
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1
XX
SQ Sequence 197 AA;
Query Match 96.1%; Score 981; DB 21; Length 197;
Best Local Similarity 96.4%; Pred. No. 4.1e-94;
Matches 190; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 1 MKKKGSVVIVGRVILNG--AVAQOTRGEGCQETSGTGRDKNQVGEVQIVSTAAQTFLA 58
DB 1 MKKKGSVVIVGRVILNGSDTAYAAQOTRGEGCQETSGTGRDKNQVGEVQIVSTATQTFLA 60
QY 59 TCINGVCMVTYVHGAGTRTIAAPKGPVIOYMTNVNDKDLVGPAPQGSRLTPTCTCGSSDLY 118
DB 61 TCINGVCMVTYVHGAGTRTIAAPKGPVIOYMTNVNDKDLVGPAPQGSRLTPTCTCGSSDLY 120
QY 119 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGPLLCPCPAGHVGIFRAAVCTRGVAK 178
DB 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGPLLCPCPAGHVGIFRAAVCTRGVAK 180
QY 179 AVDFIPVESLETTMRSP 195
DB 181 AVDFIPVESLETTMRSP 197
RESULT 5
AAB15226
ID AAB15226 standard; protein: 197 AA.
XX
AC AAB15226;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease #8.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; muteln.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.

```

DR N-PSDB; AAA73335.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Example 5; Fig 18; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polypeptide produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0  
 CC wild-type sequence.  
 XX  
 SQ Sequence 197 AA;  
 Query Match 93.9%; Score 959; DB 21; Length 197;  
 Best Local Similarity 95.4%; Pred. No. 8.2e-92;  
 Matches 188; Conservative 1; Mismatches 6; Indels 2; Gaps 1;  
 QY 1 MKKGSVIVGRIVLNG--AYAQOTRGECCOETSGTGRDNQVEGEVQIVSTAQTFLA 58  
 DB 1 MKKGSVIVGRIVLNG--AYAQOTRGECCOETSGTGRDNQVEGEVQIVSTAQTFLA 60  
 QY 59 TCINGVCVTYHAGCTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRLTPCTCGSSDLY 118  
 DB 61 TCINGVCVTYHAGCTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRLTPCTCGSSDLY 120  
 QY 119 LVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAK 178  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 QY 179 AVDFIPVESLETTMRSP 195  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 6  
 AAB15223  
 ID AAB15223 standard; protein; 197 AA.  
 XX  
 AC AAB15223;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #5.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;

DR WPI: 2000-465976/40.  
 DR N-PSDB; AAA73332.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 23; Fig 15; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polypeptide produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA;  
 Query Match 93.1%; Score 951; DB 21; Length 197;  
 Best Local Similarity 94.9%; Pred. No. 5.6e-91;  
 Matches 187; Conservative 1; Mismatches 7; Indels 2; Gaps 1;  
 QY 1 MKKGSVIVGRIVLNG--AYAQOTRGECCOETSGTGRDNQVEGEVQIVSTAQTFLA 58  
 DB 1 MKKGSVIVGRIVLNG--AYAQOTRGECCOETSGTGRDNQVEGEVQIVSTAQTFLA 60  
 QY 59 TCINGVCVTYHAGCTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRLTPCTCGSSDLY 118  
 DB 61 TSINGVLTYYHAGCTRTIASPKGPVQIOMYTNVDKDLVGNPAPQGSRLTPCTCGSSDLY 120  
 QY 119 LVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAK 178  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 QY 179 AVDFIPVESLETTMRSP 195  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 7  
 AAB15224  
 ID AAB15224 standard; protein; 197 AA.  
 XX  
 AC AAB15224;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #6.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.  
 DR N-PSDB; AAA73333.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 23; Fig 16; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 92.0%; Score 939; DB 21; Length 197;  
 Best Local Similarity 93.4%; Pred. No. 1e-89;  
 Matches 184; Conservative 3; Mismatches 8; Indels 2; Gaps 1;  
 QY 1 MKKGSVVIVGRIVLNG--AYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 58  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 60  
 QY 59 TCINGVCWTVYHGAGTRTIASPKGPVQMYTNVDKDLVGPAPQGSRLTPTCTCGSSDLY 118  
 DB 61 TSINGVLTWTVYHGAGTRTIASPKGPVQMYTNVDKDLVGPAPQGSRLTPTCTCGSSDLY 120  
 QY 119 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPGLLCPCAGHAGVIFRAAVCTRGVAK 178  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPGLLCPCAGHAGVIFRAAVCTRGVAK 180  
 QY 179 AVDFIPVESLETTMRSP 195  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 8  
 AAB15225  
 ID AAB15225 standard; protein; 197 AA.  
 XX  
 AC AAB15225;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #7.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutain.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB; AAA73334.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 23; Fig 17; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 91.0%; Score 929; DB 21; Length 197;  
 Best Local Similarity 92.9%; Pred. No. 1.1e-88;  
 Matches 183; Conservative 3; Mismatches 9; Indels 2; Gaps 1;  
 QY 1 MKKGSVVIVGRIVLNG--AYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 58  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVSTAAQTFLA 60  
 QY 59 TCINGVCWTVYHGAGTRTIASPKGPVQMYTNVDKDLVGPAPQGSRLTPTCTCGSSDLY 118  
 DB 61 TSINGVLTWTVYHGAGTRTIASPKGPVQMYTNVDKDLVGPAPQGSRLTPTCTCGSSDLY 120  
 QY 119 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPGLLCPCAGHAGVIFRAAVCTRGVAK 178  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPGLLCPCAGHAGVIFRAAVCTRGVAK 180  
 QY 179 AVDFIPVESLETTMRSP 195  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 9  
 AAY44728  
 ID AAY44728 standard; protein; 191 AA.  
 XX  
 AC AAY44728;  
 XX  
 DT 04-MAY-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 catalytic domain fusion protein-1.  
 XX  
 KW NS3 catalytic domain; NS4A peptide; NS4A-NS3 fusion construct; diagnosis;  
 KW serine protease; trypsin family; screening; anti-viral compound;  
 KW treatment; inhibitor; therapeutic.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PH Key Location/Qualifiers  
 FT Peptide 1..12  
 FT /label= NS4A\_peptide\_1  
 FT /note= "Covalently attached to amino terminus of NS3  
 FT catalytic domain"  
 FT Misc-difference 13  
 FT /note= "Wild type Proline is replaced with Lysine"  
 XX  
 PN WO200001718-A2.

```

XX PD 13-JAN-2000.
XX PF 02-JUL-1999; 99WO-US15035.
XX PR 02-JUL-1998; 98US-0091675.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Dunn BM, Bukhtiyarova M;
XX DR WPI; 2000-182103/16.
XX PT Novel polypeptide comprising hepatitis C virus NS4A and NS3 domains,
XX PT useful for screening for compounds useful for the diagnosis and
XX PT treatment of hepatitis C virus .
XX PS Claim 4; Fig 2; 30pp; English.
XX CC The present protein sequence is the fusion polypeptide, comprising the
XX CC hepatitis C virus NS4A peptidic-1 fragment, covalently attached to the
XX CC amino terminus of NS3 catalytic domain. This fusion polypeptide contains
XX CC the NS3 domain expressed in a stable, soluble form. This facilitates the
XX CC use of the polypeptide in direct screening of potential anti-viral
XX CC compounds, that are used for diagnosis and treatment of hepatitis C virus
XX CC infection. It is also used to screen for inhibitors of serine protease
XX CC activity. The polynucleotides are also useful to identify diagnostic or
XX CC therapeutic compounds and for recombinant production of the fusion
XX CC polypeptide.
XX SQ Sequence 191 AA:
Query Match 88.1%; Score 899.5; DB 21; Length 191;
Best Local Similarity 92.1%; Pred. No. 1.3e-85;
Matches 175; Conservative 6; Mismatches 8; Indels 1; Gaps 1;
QY 6 SVIVIGVILNGAYAOOTRGECCOETSGTRKNOVEGEVOIVSTAAOTELATCINGVC 65
DB 2 SVIVIGRIVLS-KYAOTRGLGCIITSLTGRKNOVEGEIQIVSTAAOTELATCINGVR 60
QY 66 WTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 125
DB 61 WTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCTCGSSDLYLVTRHAD 120
QY 126 VIPVRRRGDSRGSLSPPRISYIKGSSGGLLCPAGHANGIFRAAVCTRCVAKAVDFIPV 185
DB 121 VIPVRRRGDSRGSLSPPRISYIKGSSGGLLCPAGHANGIFRAAVCTRCVAKAVDFIPV 180
QY 186 ESLETTMRSP 195
DB 181 ENLETTMRSP 190
RESULT 10
AAR34009
ID AAR34009 standard; Protein: 2816 AA.
XX AC AAR34009;
XX AC AAR34009;
XX DT 25-MAR-2003 (updated)
XX DT 26-JUL-1993 (first entry)
XX DE HCV-1 polypeptide.
XX KW Polymerase chain reaction; PCR; amplify; primer: hepatitis C virus;
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS Hepatitis C virus.
XX PN W09306126-A1.
XX PR

```

```

PD 01-APR-1993.
XX PF 11-SEP-1992; 92WO-US07683.
XX PR 13-SEP-1991; 91US-0759575.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Weiner AJ;
XX DR WPI; 1993-117468/14.
XX PT Immuno-reactive hepatitis C virus polypeptide compans. - contg.
XX PT at least 2 sequences from the first variable domain of distinct
XX PT HCV isolates
XX PS Disclosure; Fig 9; 106pp; English.
XX CC This sequence represents the entire hepatitis C virus polypeptide.
XX CC HCV is a member of the flavivirus family and appears to encode a basic
XX CC polypeptide domain ("C") at the N-terminal of the viral polypeptide,
XX CC followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
XX CC nonstructural genes NS2 through NS5. See also AAQ39134-48, AAR33982-
XX CC 4008 and AAR36088-89.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 2816 AA:
Query Match 88.0%; Score 898.5; DB 14; Length 2816;
Best Local Similarity 85.8%; Pred. No. 6.1e-84;
Matches 175; Conservative 9; Mismatches 9; Indels 11; Gaps 2;
QY 3 KKGSVIVIG---RIVLNG-----AYAOOTRGECCOETSGTRKNOVEGEVOIVST 51
DB 1005 RRGREILLGPADGVMVMKGRLLAPITAYAOOTRGLGCIITSLTGRKNOVEGEVOIVST 1064
QY 52 AAOTFLATCINGVCMTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCT 111
DB 1065 AAOTFLATCINGVCMTVYHGAGTRTIA SPKGPVIQMTNVNDKDLVGPAPQGSRLTPCT 1124
QY 112 CGSSDLYLVTRHADVIPVRRRGDSRGSLSPPRISYIKGSSGGLLCPAGHANGIFRAAV 171
DB 1125 CGSSDLYLVTRHADVIPVRRRGDSRGSLSPPRISYIKGSSGGLLCPAGHANGIFRAAV 1184
QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
RESULT 11
AAR40120
ID AAR40120 standard; Protein: 3011 AA.
XX AC AAR40120;
XX AC AAR40120;
XX DT 25-MAR-2003 (updated)
XX DT 27-JAN-1994 (first entry)
XX DE HCV genomic amino acid sequence isolated from infected human LG.
XX KW Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
XX KW human growth hormone; HGH; secretion signal; fusion protein;
XX KW vaccine.
XX OS Hepatitis C Virus.
XX PN W09315193-A1.
XX PR 05-AUG-1993.
XX PF 29-JAN-1993; 93WO-US00907.
XX PR 31-JAN-1992; 92US-0830024.

```

XX (ABBO ) ABBOTT LAB.  
 PA Bode SL, Cassey JM, Desai SM, Davaro SG, Fraill DE;  
 PI Yamaguchi J, Zock BJ;  
 XX WPI; 1993-258673/32.  
 DR  
 XX  
 XX New plasmid pHCV-162 is a mammalian expression systems for HCV D  
 PT proteins - useful for diagnosing HCV infection and as vaccines  
 PT for preventing HCV infection  
 XX  
 XX  
 XX Example 1: Page 39-49; 100pp; English.  
 PS  
 XX RNA was isolated from the plasma of a HCV seropositive human  
 CC (designated "LG") and cDNA was prepared from it. The cDNA was  
 CC PCR amplified using specific primers with sequences based  
 CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further  
 CC amplification using nested primers resulted in 7 adjacent HCV DNA  
 CC fragments which could be assembled into a full-length sequence. The  
 CC DNA sequence was determined and translated into the genomic amino  
 CC acid sequence. Comparison of the LG genomic amino acid sequence  
 CC with that from HCV-1 showed 134 amino acid differences.  
 CC (Updated on 23-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 3011 AA;

Query Match 87.7%; Score 895.5; DB 14; Length 3011;  
 Best Local Similarity 85.8%; Pred. No. 1.4e-83;  
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;  
 OY 3 KGSVVIVG---RIVLNG-----AYAQTRGEGCGQETSGTRGNQVEGEVQIVST 51  
 DB 1005 RRGREILGPADGMVSKGWRLLAPITAYAQTRGILGCLITSLGRDNQVEGEVQIVST 1054  
 OY 52 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVQIOMYTNVDKDLVGPAPQGSRLTPCT 111  
 DB 1065 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVQIOMYTNVDKDLVGPAPQGSRLTPCT 1124  
 OY 112 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 171  
 DB 1125 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 1184  
 OY 172 CTRGVAKAVDFIPVESLEITMRSP 195  
 DB 1185 CTRGVAKAVDFIPVESLEITMRSP 1208

RESULT 12  
 AAP92041  
 ID AAP92041 standard; protein; 1766 AA.  
 XX  
 AC AAP92041;

XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)

DE Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones  
 DE 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,  
 DE 33f, 33g and 39c.

XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HABH)

XX Hepatitis C virus.

XX EP318216-A.

XX 31-MAY-1989.

XX 18-NOV-1988; 88EP-0310922.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Houghton M, Choo QL, Kuo G;  
 XX WPI; 1989-159274/22.  
 DR N-PSDB; AAN92097.  
 XX  
 XX Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 PT  
 XX Claim 13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.  
 XX  
 CC It is the sequence encoded in the open reading frame of hepatitis C virus  
 CC cDNA inserts in clones 14i, m 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,  
 CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could  
 CC be used in immunosay reagents and vaccines and to generate antibodies  
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,  
 CC non-B hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 1766 AA;

Query Match 87.6%; Score 894.5; DB 10; Length 1766;  
 Best Local Similarity 85.8%; Pred. No. 8.6e-84;  
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;  
 OY 3 KGSVVIVG---RIVLNG-----AYAQTRGEGCGQETSGTRGNQVEGEVQIVST 51  
 DB 289 RRGREILGPADGMVSKGWRLLAPITAYAQTRGILGCLITSLGRDNQVEGEVQIVST 348  
 OY 52 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVQIOMYTNVDKDLVGPAPQGSRLTPCT 111  
 DB 349 AAQTFLATCINGCVTVYHGAGTRTITASPKGPVQIOMYTNVDKDLVGPAPQGSRLTPCT 408  
 OY 112 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 171  
 DB 409 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 468  
 OY 172 CTRGVAKAVDFIPVESLEITMRSP 195  
 DB 469 CTRGVAKAVDFIPVENLEITMRSP 492

RESULT 13

AAP90158  
 ID AAP90158 standard; protein; 1786 AA.

XX  
 AC AAP90158;

XX 25-MAR-2003 (updated)

DT 10-NOV-1989 (first entry)

XX Protein sequence of hepatitis c virus composite cdna.

DE Hepatitis C virus; vaccine.

XX Pan troglodytes.

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

PR 30-DEC-1987; 87US-0139886.

PR 26-FEB-1988; 88US-0161072.

```

PR 26-OCT-1988; 88US-0263584.
XX (CHIR ) CHIRON CORPORATION.
PA Houghton M, Choo QL, Kuo G;
PI WPI; 1989-215054/30.
XX N-PSDB: AAN90327.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment
PT of infection.
XX Disclosure: fig 26; 30pp; English.
PS The sequence is encoded by the composite cDNA of AAN90327. These
CC antigens react with antibodies in patients with non-B hepatitis
CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC antibodies for immunoassay or treatment, or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 1786 AA;
SQ
Query Match 87.68; Score 894.5; DB 10; Length 1786;
Best Local Similarity 85.8%; Pred. No. 8.7e-84;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
QY 3 KKGSVWIVG---RIVLNG-----AYAOOTRGECCOETSTGTGRKNOVEGEVIVST 51
DB 289 RRGREILLGPDAGWVSKGWRLLAPITAYAOOTRGLGCIITSLTGRKNOVEGEVIVST 348
QY 52 AAOTFLATCINGVCWTVYHGAGTNTIASPKGPVIQMYTNVDKDLVGPAPQGSRSITPCT 111
DB 349 AAOTFLATCINGVCWTVYHGAGTNTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCT 408
QY 112 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 171
DB 409 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 468
QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 469 CTRGVAKAVDFIPVENLETTMRSP 492
RESULT 14
AAP90164
ID AAP90164 standard; protein: 2261 AA.
XX AAP90164;
XX 25-MAR-2003 (updated)
DT 01-NOV-1989 (first entry)
XX Peptide encoded by composite hepatitis C virus cDNA.
DE Hepatitis C virus; clone 12f; clone 15e; probe: vaccine.
XX Pan troglodytes.
XX GB2212511-A.
XX 26-JUL-1989.
XX 18-NOV-1988; 88GB-0027024.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 26-OCT-1988; 88US-0263584.
XX (CHIR ) CHIRON CORPORATION
PA Houghton M, Choo QL, Kuo G;
PI WPI; 1989-215054/30.
XX N-PSDB: AAN90331.
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT polypeptide(s) and antibodies for diagnosis, prevention and
PT treatment of infection.
XX Disclosure: fig 32; 235pp; English.
PS The sequence is the peptide encoded by the composite hepatitis C
CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose
CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,
CC or to produce vaccines.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX Sequence 2261 AA;
SQ
Query Match 87.6%; Score 894.5; DB 10; Length 2261;
Best Local Similarity 85.8%; Pred. No. 1.2e-83;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;
QY 3 KKGSVWIVG---RIVLNG-----AYAOOTRGECCOETSTGTGRKNOVEGEVIVST 51
DB 380 RRGREILLGPDAGWVSKGWRLLAPITAYAOOTRGLGCIITSLTGRKNOVEGEVIVST 439
QY 52 AAOTFLATCINGVCWTVYHGAGTNTIASPKGPVIQMYTNVDKDLVGPAPQGSRSITPCT 111
DB 440 AAOTFLATCINGVCWTVYHGAGTNTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCT 499
QY 112 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 171
DB 500 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 559
QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 560 CTRGVAKAVDFIPVENLETTMRSP 583

```

```

RESULT 15
AAP92047
ID AAP92047 standard; protein: 2301 AA.
XX AAP92047;
XX 25-MAR-2003 (updated)
DT 02-MAR-1990 (first entry)
XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE 12f through 15e.
XX Hepatitis C virus (HCV); non-A, non-B hepatitis (NANBH).
XX Hepatitis C virus.
XX EP318216-A.
XX 31-MAY-1989.
XX 18-NOV-1988; 88EP-0310922.
XX 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
XX (CHIR ) CHIRON CORP.
XX Houghton M, Choo QL, Kuo G;
XX WPI; 1989-159274/22.

```

DR N-PSDB; AAN92103.  
XX  
PT Purified hepatitis C virus  
PT - and associated nucleic acids and polypeptide(s)  
XX  
PS Claim 13; Figure 32-1 - 32-7; 139 pp; English.  
XX  
CC It is the sequence encoded in the open reading frame of hepatitis C virus  
CC (HCV) cDNA inserts in clones 12f through 15c. It is antigenic and could  
CC be used in immunosay reagents and vaccines and to generate antibodies  
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,  
CC non-B hepatitis.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX  
SQ Sequence 2301 AA:  
  
Query Match 87.6%; Score 894.5; DB 10; Length 2301;  
Best Local Similarity 85.8%; Pred. No. 1.2e-83;  
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;  
  
QY 3 KKGSVVVG---RVLNG-----AVAQOTRGECCQFTSOTGRDNQVEGEVQIVST 51  
DB 360 RRGREILGPADGMVSKGWRLAPITAYAQOTRGLLGCITISLGRDNQVEGEVQIVST 439  
  
QY 52 AAQTFLATCINGVCWTVYHGAGIRTIASPKGPVIOMYTNVDKDLVGPAPQGSRSLEPT 111  
DB 440 AAQTFLATCINGVCWTVYHGAGIRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSLEPT 499  
  
QY 112 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPAGHAVGIFRAAV 171  
DB 500 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPAGHAVGIFRAAV 559  
  
QY 172 CTRGVAKAVDFIPVESLETMRSP 195  
DB 560 CTRGVAKAVDFIPVENLETMRSP 583

Search completed: August 30, 2003, 19:12:22  
Job time : 47.1697 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 16.0488 Seconds  
(without alignments)  
1168.492 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKKGSVIVGRIVLNGAYA.....VAKAVDFIPVESLETIMKSP 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Length	DR	ID	Description
1	894.5	87.6	3011	1	GNMVC3	genome polyprotein
2	893.5	87.5	3011	1	S40770	genome polyprotein
3	879.5	86.1	3011	1	GNMVC3	genome polyprotein
4	847	83.0	3010	1	GNMVC3	genome polyprotein
5	842	82.5	3010	1	GNMVC3	genome polyprotein
6	838	82.1	3010	1	GNMVC3	genome polyprotein
7	837	82.0	3010	1	A45573	genome polyprotein
8	821	80.4	3010	1	S18030	genome polyprotein
9	761	74.5	3014	1	JC5620	genome polyprotein
10	670.5	65.7	3033	1	JQ1303	genome polyprotein
11	668.5	65.5	3033	1	GNMVC3	genome polyprotein
12	260	25.5	2970	2	T08839	polyprotein - marm
13	258	25.3	3005	2	T08841	polyprotein - dour
14	85.5	8.4	495	2	B71360	hypothetical prote
15	85.5	8.4	1334	2	AB1775	hypothetical prote
16	82	8.0	452	2	J39383	angio-associated m
17	80.5	7.9	590	2	B81104	nitrate/nitrite se
18	80.5	7.9	590	2	C81911	nitrate/nitrite se
19	79.5	7.8	642	1	VCMVFG	env polyprotein -
20	79	7.7	259	1	IOH01	insulin-like growt
21	79	7.7	354	2	T49806	hypothetical prote
22	79	7.7	404	2	A46165	envelope surface g
23	78.5	7.7	209	2	H83144	probable aromatic
24	78.5	7.7	981	2	T18234	beta transducin ho
25	77.5	7.6	398	2	B71284	probable periplasm
26	77.5	7.6	2663	1	S28261	centromere protein
27	77.5	7.6	3507	2	T34513	hypothetical prote
28	76.5	7.5	270	2	T06118	hypothetical prote
29	76.5	7.5	393	2	E95261	serine proteinase

30 76.5 7.5 397 2 B98127 serine proteinase  
31 76 7.4 140 2 C72705 hypothetical prote  
32 76 7.4 574 2 A84782 hypothetical prote  
33 75.5 7.4 859 2 T35785 probable beta-gluc  
34 75.5 7.4 882 2 S41034 hypothetical prote  
35 75 7.3 639 1 VCMVSA env polyprotein pr  
36 75 7.3 1293 2 T30871 orsellinic acid sy  
37 74.5 7.3 415 2 S70401 zona pellucida gly  
38 74.5 7.3 492 2 AH1030 probable exported  
39 74.5 7.3 755 2 S23441 hypothetical prote  
40 74.5 7.3 1176 2 T18042 ice nucleation pro  
41 74 7.2 239 2 H89966 serine proteinase  
42 74 7.2 239 2 G87265 conserved hypothet  
43 74 7.2 603 1 VCFVER env polyprotein -  
44 73.5 7.2 317 2 S76618 hypothetical prote  
45 73.5 7.2 566 2 H84203 phosphate ABC tran

ALIGNMENTS

genome polyprotein - hepatitis C virus (strain HCV-1)  
M:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HCV-1)  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence:revision 30-Sep-1992 #text\_change 19-Jan-2001  
C:Accession: A39166; PQ0403; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; MUID:91172826; PMID:1848704  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peuchere, J.F.; Follett, E.; Yap,  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0403  
A:Molecule type: genomic RNA  
A:Residues: 1577-1633 <CHA>  
A:Cross-references: DDBJ:D10128  
A:Experimental source: isolates E-b16  
A:Accession: PQ0404  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1577-1633 <CH2>  
A:Experimental source: isolates E-b17  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 87.6% ; Score 894.5 ; DB 1 ; Length 3011 ;

Best Local Similarity 85.8% ; Pred. No. 1.7e-74 ;

Matches 175 ; Conservative 8 ; Mismatches 10 ; Indels 11 ; Gaps 2 ;

QY 3 KKGSVIVG---RIVLNG-----AYAQTRGECQETSOTGRKNQVEGVQIVST 51  
::: ::: :| | ||||| | | ||||| |||||



```
Db 1005 RRCKEILLGPADQWMSKGVKLLAPITAYAAQITGLGLGCIITSLTGRDKNQVEGEVQIVST 1064
Qy 52 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 111
Db 1065 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 2
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis M #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 87.5% Score 893.5; DB 1: Length 3011;
Best Local Similarity 85.8%; Pred. No. 2,1e-74;
Matches 175; Conservative 7; Mismatches 11; Indels 11; Gaps 2;

Qy 3 KKGSVTVG---RIVLNG-----AYAOOTRGECCOETISOTGRDKNQVEGEVQIVST 51
Db 1005 RKGEIILLGPADGMVSKGWLLAPITAYAAOQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
Qy 52 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 111
Db 1065 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208
```

```
RESULT 3
GNVWCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedes, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A41546; MUID:92052256; PMID:1658600
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEP>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis M #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 86.1% Score 879.5; DB 1: Length 3011;
Best Local Similarity 83.8%; Pred. No. 4,1e-73;
Matches 171; Conservative 9; Mismatches 13; Indels 11; Gaps 2;

Qy 3 KKGSVTVG---RIVLNG-----AYAOOTRGECCOETISOTGRDKNQVEGEVQIVST 51
Db 1005 RRGOEILLGPADGMVSKGWLLAPITAYAAOQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
Qy 52 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 111
Db 1065 AAOTFLATCINGVCTVYHAGAGTTRTASPKGPVIQWYINVKDVLVGVWPAQGSRLTPTCT 1124
Qy 112 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 171
Db 1125 CGSSDLVLTVRHADVIVPVRKRGDSRGSLSPRPISYLKSSGGPGLCPAGHAGVIFRAAV 1184
Qy 172 CTRGVAKAVDFIPVESLETTMRSP 195
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
GNVWCH
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis M; hepatitis C virus (strain Taiwan)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
```

Matches	154;	Conservative	14;	Mismatches	10;	Indels	0;	Gaps	0;
Qy	18	AYAAQTGEEGCQET	SOTGRDKNOVEGOIVSTAATFLATCINGVCWTVYHGAGTRTI	77					
Dd	1031	AYSQOTRGLLCIITSLTGRDKNOVEGOVVS	TATQSFLATCNGVCWTVYHAGSKTL	1090					
Qy	78	ASPKPGVIQMTYNVDKDLVGMPAPOGSRS	LTPCTCGSSDLYLVRHADVIPRRRGDSRG	137					
Dd	1091	AAPKGPIQMTYNVDQLVGMPKPGARSLT	PCTCGSSDLYLVRHADVIPRRRGDSRG	1150					
Qy	138	SLLSPRPISYLKSGGGPLLCPAGHAVCI	FRAAVCTRCVAKAVDFIPVESLETTMRSP	195					
Dd	1151	SLLSPRPVSYLKSGGGPLLCPPGHAYGI	FRAAVCTRCVAKAVDFIPVESMETMTRSP	1208					
<b>RESULT 6</b>									
<b>GNVWCJ</b>									
genome polyprotein - hepatitis C virus (strain J)									
N:	Contains:	capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5							
C:	Species:	hepatitis C virus							
C:	Date:	30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001							
C:	Accession:	A39253; PS0086							
R:	Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shirogane, K.	Proc. Natl. Acad. Sci. U.S.A.	87,	9524-9528,	1990				
A:	Title:	Molecular cloning of the human Hepatitis C virus genome from Japanese patients							
A:	Reference number:	A39253; PMID:91088550; PMID:2175903							
A:	Accession:	A39253							
A:	Molecule type:	genomic RNA							
A:	Residues:	1-3010 <KAT>							
A:	Cross-references:	GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611							
R:	Kato, N.; Ohkoshi, S.; Shimotohno, K.	Proc. Jpn. Acad.	65B,	219-223,	1989				
A:	Title:	Japanese isolates of the non-A, non-B hepatitis viral genome show sequence homology with the non-A, non-B hepatitis viral genome							
A:	Reference number:	PS0085							
A:	Accession:	PS0086							
A:	Molecule type:	genomic RNA							
A:	Residues:	2650-2707 <KA2>							
A:	Experimental source:	Japanese isolate							
C:	Comment:	The cleavage sites of this polyprotein have not been determined.							
C:	Superfamily:	hepatitis C virus genome polyprotein							
C:	Keywords:	ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; signal peptide							
F:	2-115/Product:	capsid protein C #status predicted <CP>							
F:	116-191/Product:	envelope protein M #status predicted <EPM>							
F:	192-389/Product:	major envelope protein E #status predicted <ME>							
F:	390-729/Product:	nonstructural protein NS1 #status predicted <NS1>							
F:	730-1006/Product:	nonstructural protein NS2 #status predicted <NS2>							
F:	1007-1615/Product:	hepacivirus #status predicted <NS3>							
F:	1230-12317/Region:	nucleotide-binding motif A (P-loop)							
F:	1312-1317/Region:	nucleotide-binding motif B							
F:	1316-1319/Region:	DXH motif							
F:	1616-1662/Product:	nonstructural protein NS4a #status predicted <N4a>							
F:	1863-2013/Product:	nonstructural protein NS4b #status predicted <N4b>							
F:	2014-3010/Product:	nonstructural protein NS5 #status predicted <NS5>							
F:	196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2172,2173,2174,2175,2176,2177,2178,2179,2180,2181,2182,2183,2184,2185,2186,2187,2188,2189,2190,2191,2192,2193,2194,2195,2196,2197,2198,2199,2200,2201,2202,2203,2204,2205,2206,2207,2208,2209,2210,2211,2212,2213,2214,2215,2216,2217,2218,2219,2220,2221,2222,2223,2224,2225,2226,2227,2228,2229,2230,2231,2232,2233,2234,2235,2236,2237,2238,2239,2240,2241,2242,2243,2244,2245,2246,2247,2248,2249,2250,2251,2252,2253,2254,2255,2256,2257,2258,2259,2260,2261,2262,2263,2264,2265,2266,2267,2268,2269,2270,2271,2272,2273,2274,2275,2276,2277,2278,2279,2280,2281,2282,2283,2284,2285,2286,2287,2288,2289,2290,2291,2292,2293,2294,2295,2296,2297,2298,2299,2300,2301,2302,2303,2304,2305,2306,2307,2308,2309,2310,2311,2312,2313,2314,2315,2316,2317,2318,2319,2320,2321,2322,2323,2324,2325,2326,2327,2328,2329,2330,2331,2332,2333,2334,2335,2336,2337,2338,2339,2340,2341,2342,2343,2344,2345,2346,2347,2348,2349,2350,2351,2352,2353,2354,2355,2356,2357,2358,2359,2360,2361,2362,2363,2364,2365,2366,2367,2368,2369,2370,2371,2372,2373,2374,2375,2376,2377,2378,2379,2380,2381,2382,2383,2384,2385,2386,2387,2388,2389,2390,2391,2392,2393,2394,2395,2396,2397,2398,2399,2400,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2								

## RESULT 7

A45573  
genome polyprotein - hepatitis C virus (strain J7)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:132-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
Query Match 82.0%; Score 837; DB 1; Length 3010;  
Best Local Similarity 86.5%; Pred. No. 3 7e-69;  
Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77  
DB 1031 AYAAQTRGLLGLGIVTSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090  
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 137  
DB 1091 AGPKGPITQMTNVQDVLGVWPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 1150  
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195  
DB 1151 SLLSPRPVSYLGSSGGPILCPGSHAGVIFRAAVCTRGVAKAVDFIPVESMETTMRSP 1208

## RESULT 8

S18030  
genome polyprotein - hepatitis C virus (isolate JK1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:132-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEXH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
Query Match 82.0%; Score 837; DB 1; Length 3010;  
Best Local Similarity 86.5%; Pred. No. 3 7e-69;  
Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77  
DB 1031 AYAAQTRGLLGLGIVTSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090  
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 137  
DB 1091 AGPKGPITQMTNVQDVLGVWPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 1150  
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195  
DB 1151 SLLSPRPVSYLGSSGGPILCPGSHAGVIFRAAVCTRGVAKAVDFIPVESMETTMRSP 1208

## A:Accession: S33570

A:Molecule type: genomic RNA  
A:Residues: 1-547,'T','V',549-621,'V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 3  
as Trp, and TTC for residue 771 as Ser  
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:132-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196-209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate  
Query Match 80.4%; Score 821; DB 1; Length 3010;  
Best Local Similarity 85.4%; Pred. No. 1.1e-67;  
Matches 152; Conservative 13; Mismatches 13; Indels 0; Gaps 0;  
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77  
DB 1031 AYAAQTRGLLGLGIVTSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090  
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 137  
DB 1091 AGPKGPITQMTNVQDVLGVWPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 1150  
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195  
DB 1151 SLLSPRPVSYLGSSGGPILCPGSHAGVIFRAAVCTRGVAKAVDFIPVESMETTMRSP 1208

## RESULT 9

JC5620  
genome polyprotein - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:132-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEXH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
Query Match 80.4%; Score 821; DB 1; Length 3010;  
Best Local Similarity 85.4%; Pred. No. 1.1e-67;  
Matches 152; Conservative 13; Mismatches 13; Indels 0; Gaps 0;  
QY 18 AYAAQTRGEGCGEQTSGTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77  
DB 1031 AYAAQTRGLLGLGIVTSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGSKTL 1090  
QY 78 ASPKGPVIQMTNVKDLVGVPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 137  
DB 1091 AGPKGPITQMTNVQDVLGVWPAQGSRLTPTCTGSSDLVLTTRHADVTPVRRGDSRG 1150  
QY 138 SLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195  
DB 1151 SLLSPRPVSYLGSSGGPILCPGSHAGVIFRAAVCTRGVAKAVDFIPVESMETTMRSP 1208

F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 74.5%; Score 761; DB 1; Length 3014;  
Best Local Similarity 77.5%; Pred. No. 4.3e-62;  
Matches 138; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

QY 18 AYAQOTRGECCQTSOTGRDNQVEGEVQIVSTAQTFLATCINGVCVWTVYHGAGTRTI 77  
DB 1032 AYAQOTRGLCAIVLSLTGRDKNEAGEVQFLSTATQTFICICINGVWTVLPHGAGSKTL 1091

QY 78 ASPKGPVQMTYNDKDLVGHWPAPQGSRLTPCTCGSSDLYLVRHADVPVRRRGDSRG 137  
DB 1092 AGPKGPVQMTYNDKDLVGHWPSPGKSLRCTCGSADLYLVRHADVPVRRRGDTTRA 1151

QY 138 SLLSPRISYLGSGGPGLLCPAGHAGVIFRAAICTRGAKAVDFIPVESLETTMRSP 195  
DB 1152 SLLSPRISYLGSGGPGIMCPSGHVGVFRAAICTRGAKAVEFPVENLETTMRSP 1209

RESULT 10  
J01303  
genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C:Accession: J01303  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Mizuka, H.; Machida, A.; Miyakawa, Y.  
J. Gen. Virol. 72, 2697-2704, 1991  
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
A:Reference number: J01303; MUID:92044440; PMID:1658196  
A:Accession: J01303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651  
A:Experimental source: isolate HC-J6 from a Japanese individual  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmembrane  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:1011-1619/Product: hepatitis C virus NS2 #status predicted <NS2>  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4A>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,26

Query Match 65.7%; Score 670.5; DB 1; Length 3033;  
Best Local Similarity 62.1%; Pred. No. 1.1e-53;  
Matches 126; Conservative 27; Mismatches 33; Indels 17; Gaps 1;

QY 10 VGRIVLNG-----AYAQOTRGECCQTSOTGRDNQVEGEVQIVSTA 52  
DB 1010 LGREVLGPGADYTSKGNLSLLAPITAYAQOTRGLGTVVSVMTGRDTEQAGEIQVLSLV 1069

QY 53 AOTFLATCINGVCVWTVYHGAGTRTIASPKGPVQIOMYTNVDKDLVGHWPAPQGSRLTPCTC 112  
DB 1070 TQFLGTISGVLTIVYHGAGNKTLAGSGRPVQIOMYTNVDKDLVGHWPAPQGSRLTPCTC 1129

QY 113 GSSDLYLVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAA 172  
DB 1130 GAVDLYLVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAA 1189

QY 173 TRGKAVDFIPVESLETTMRSP 195  
DB 1190 SRGKAVSDFIPVESLTVTRSP 1212

RESULT 11  
GNMYJ8  
genome polyprotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C:Species: hepatitis C virus  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40250; PQ0397; PQ0559  
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Mizuka, H.; Tanaka, T.; Fukuda, Y.  
Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to  
A:Reference number: A40250; MUID:92230232; PMID:1314459  
A:Accession: A40250  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0397  
A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D10134  
A:Experimental source: isolate E-bl2  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotoh  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: PQ0554; MUID:92068204; PMID:1720309  
A:Accession: PQ0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <XAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis C virus NS3 #status predicted <NS3>  
F:1324-1241/Region: nucleotide-binding motif A (p-loop)  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4A>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

Query Match 65.5%; Score 668.5; DB 1; Length 3033;  
Best Local Similarity 62.6%; Pred. No. 1.8e-53;  
Matches 127; Conservative 25; Mismatches 34; Indels 17; Gaps 1;

QY 10 VGRIVLNG-----AYAQOTRGECCQTSOTGRDNQVEGEVQIVSTA 52  
DB 1010 LGREVLGPGADYTSKGNLSLLAPITAYAQOTRGLGTVVSVMTGRDTEQAGEIQVLSLV 1069

QY 53 AOTFLATCINGVCVWTVYHGAGTRTIASPKGPVQIOMYTNVDKDLVGHWPAPQGSRLTPCTC 112  
DB 1070 TQFLGTISGVLTIVYHGAGNKTLAGSGRPVQIOMYTNVDKDLVGHWPAPQGSRLTPCTC 1129

QY 113 GSSDLYLVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAA 172  
DB 1130 GAVDLYLVRHADVPVRRRGDSRLSPRISYLGSGGPGLLCPAGHAGVIFRAA 1189

QY 173 TRGKAVDFIPVESLETTMRSP 195  
DB 1490 ARGVAKSIDFIPVESLTVTRTP 1212

RESULT 12  
T08839  
polyprotein - marmoset hepatitis GB virus A

[illegible]

Db	738	KGILQSLKIYDELVSVMGYPQTIIVVEMARENQTTGCKGNNSRPRYKSLEKAIKEFGSQ	797
Qy	60	CI-----NGVCWIVYHGAGTRTIAAPKGPVIQMTYNVDKDL-----VGWPAP	101
Db	798	ILKEHPTDNQELRNNRLYYLQNGK-----DMYTGODLDIHNLNSYDIDHIYP	846
Qy	102	QGSMSLTPCTCGSSDLYLVTRHA-----DVIP---VRRRGD-----SRGSIISPRPIS	146
Db	847	QSF-----ITDNSTIDNIVLTSSAGNREKGGDDVPPLEIVRKRKVFWEKLYOGNLMKRKFD	901
Qy	147	YL-KGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAV	180
Db	902	YLTRAERG--LTHADRARFIHROLVETROITKNV	934

Search completed: August 30, 2003, 19:20:26  
Job time : 18.0488 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.65768 seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKKGSVIVGRIVLNGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	894.5	87.6	3011	1 POLG_HCV1	P26664 h genome po
2	879.5	86.1	3011	1 POLG_HCVH	P27958 h genome po
3	847	83.0	3010	1 POLG_HCVTW	P29846 h genome po
4	842	82.5	3010	1 POLG_HCVBK	P26663 h genome po
5	838	82.1	3010	1 POLG_HCVJA	P26662 h genome po
6	837	82.0	3010	1 POLG_HCVUT	Q00269 h genome po
7	670.5	65.7	3033	1 POLG_HCVJ6	P26660 h genome po
8	668.5	65.5	3033	1 POLG_HCVJ8	P26661 h genome po
9	85.5	8.4	485	1 Y136_IIEPA	Q83172 treponema p
10	82.5	8.1	321	1 HHOA_ARATH	Q9sel7 arabidopsis
11	82	8.0	452	1 AAMP_HUMAN	Q13685 homo sapien
12	80.5	7.9	437	1 DEGL_ARATH	Q22609 arabidopsis
13	79	7.7	259	1 IRL1_HUMAN	P08833 homo sapien
14	78.5	7.7	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
15	77.5	7.6	642	1 ENV_FLVGL	P08359 feline leuk
16	77.5	7.6	2663	1 CENE_HUMAN	Q02224 homo sapien
17	75.5	7.4	861	1 P058_CAEEL	P34552 caenorhabdi
18	75	7.3	639	1 ENV_FLVSA	P06752 feline leuk
19	74.5	7.3	415	1 2P1_RABIT	P48833 oryctolagus
20	74.5	7.3	776	1 H1PF_AZOVI	P40596 acrobacter
21	74	7.2	402	1 RAGE_RAT	Q63495 rattus norv
22	74	7.2	603	1 ENV_RSVP	P03396 rous sarcom
23	73.5	7.2	263	1 GRAK_MOUSE	Q35205 mus musculu
24	73.5	7.2	436	1 ENV_FLVCS	Q02077 feline leuk
25	73.5	7.2	661	1 INVB_DAUCA	P80065 daucus caro
26	73.5	7.2	1165	1 POL_GALV	P21414 gibbon ape
27	73	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
28	73	7.1	645	1 ENV_FSVSM	P21445 feline sarc
29	73	7.1	676	1 ENV_MLVFP	P26803 friend muri
30	73	7.1	1705	1 PTPO_MOUSE	P70289 mus musculu
31	73	7.1	3414	1 POLG_LANVT	P29837 l genome po
32	72.5	7.1	257	1 GRAAL_HUMAN	P51124 homo sapien
33	72	7.1	659	1 VST2_HEVME	Q03500 hepatitis e

RESULT 1  
POLG\_HCV1 STANDARD: PRT; 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate 1) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,  
RA Bradley D.W., Kuo G., Houghton M.,  
RA "Genetic organization and diversity of the hepatitis C virus.";  
Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
RL CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
precursor polyprotein, commonly with Asp or Glu in the P6  
position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA)(N).  
CC -1- SURUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M62321; AAA45676.1; -  
DR PIR; A39166; GNMVC3.  
DR PDB; 1A1V; 16-FEB-99.  
DR PDB; 1HEI; 25-NOV-98.  
DR MEROPS; S29.001; -  
DR MEROPS; U39.001; -  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002522; HCV\_capsid.

P26804 friend muri  
P39061 mus musculu  
P11033 mus musculu  
Q90627 gallus gall  
P10205 herpes simp  
P09855 herpes simp  
P97608 rattus norv  
P03221 epstein-bar  
P11261 feline leuk  
P54748 rattus norv  
P12641 bovine herp  
P18614 rattus norv

ALIGNMENTS

34 72 7.1 676 1 ENV\_MLVFF  
35 72 7.1 1527 1 CAIH\_MOUSE  
36 71.5 7.0 248 1 GRAD\_MOUSE  
37 71.5 7.0 248 1 TRY1\_CHICK  
38 71.5 7.0 535 1 UL21\_HSV11  
39 71.5 7.0 535 1 UL21\_HSV11  
40 71.5 7.0 1288 1 OPLA\_RAT  
41 71 7.0 336 1 UL16\_ERV  
42 71 7.0 662 1 ENV\_FLVLB  
43 71 7.0 844 1 CNM4\_RAT  
44 71 7.0 917 1 VGLB\_HSVB2  
45 71 7.0 1180 1 ITAL\_RAT

DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002888; HCV\_NS4c.  
 DR InterPro: IPR002186; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR Pfam: PF0186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT CHAIN 3011 369 POTENTIAL.  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1319 1319 DECH\_BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327197 MW; 63F8C9447FCE5AF9 CRC64;

Query Match 87.6%; Score 894.5; DB 1; Length 3011;  
 Best Local Similarity 85.8%; Pred. No. 2,2e-76;  
 Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;  
 3 KKGWVWG---RIVLNG-----AYAOQTREGSGCOETSGTGRKKNQVGEVQIVST 51

Db 1005 RRGREILLGADGWVSKGWRLLAPITATYAQQRGLGCIITSLTGRKNQVGEVQIVST 1064  
 Qy 52 AAQTFLATCINGVCWTVYHGAGTRTIASPKGVIOYMTNVKDLVGPAPQCSRLTPCT 111  
 Db 1065 AAOTELATCINGVCWTVYHGAGTRTIASPKGVIOYMTNVQDLVGPAPQCSRLTPCT 1124  
 Qy 112 CGSSDLYLVTRHADVTPVRRGDSRGSLSPRPISYLYKSGSGGPLLCCPAGHAGVIFRAV 171  
 Db 1125 CGSSDLYLVTRHADVTPVRRGDSRGSLSPRPISYLYKSGSGGPLLCCPAGHAGVIFRAV 1184  
 Qy 172 CTRGKAVKAVDFIPVESLETTMRSP 195  
 Db 1185 CTRGKAVKAVDFIPVENLETTMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11108;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE-92052256; PubMed-1658800;  
 RX Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT \*Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RP MEDLINE-9731322; PubMed-9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT \*Structure of the hepatitis C virus RNA helicase domain.\*;  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RP MEDLINE-98154321; PubMed-9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 Murcko M.A., Lin C., Caron P.R.;  
 RT \*Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.\*;  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.



CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M67463; AAA45534.1; .  
 CC PIR: A36814; GNMVCH.  
 CC PDB: 1HEI; 25-NOV-98.  
 CC PDB: 1AIV; 16-FEB-99.  
 CC PDB: 1A1R; 17-JUN-98.  
 CC MEROPS: S29.001; .  
 CC MEROPS: U39.001; .  
 CC TRANSFAC: T04155; .  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV capsid.  
 CC InterPro: IPR002521; HCV core.  
 CC InterPro: IPR002519; HCV env.  
 CC InterPro: IPR002531; HCV NS1.  
 CC InterPro: IPR002518; HCV NS2.  
 CC InterPro: IPR004109; HCV NS3.  
 CC InterPro: IPR000745; HCV NS4a.  
 CC InterPro: IPR001490; HCV NS4b.  
 CC InterPro: IPR002868; HCV NS5a.  
 CC InterPro: IPR002166; HCV RdRp.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro: IPR007094; RNA\_pol\_PSVir.  
 CC Pfam: PF01543; HCV capsid; 1.  
 CC Pfam: PF01542; HCV core; 1.  
 CC Pfam: PF01539; HCV env; 1.  
 CC Pfam: PF01560; HCV NS1; 1.  
 CC Pfam: PF01538; HCV NS2; 1.  
 CC Pfam: PF02907; HCV NS3; 1.  
 CC Pfam: PF01006; HCV NS4a; 1.  
 CC Pfam: PF01001; HCV NS4b; 1.  
 CC Pfam: PF01506; HCV NS5a; 1.  
 CC Pfam: PF00271; helicase\_C; 1.  
 CC Pfam: PF00998; Viral\_RdRp; 1.  
 CC ProDom: PD186062; HCV NS1; 1.  
 CC SMART: SM00487; DEXDc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3d-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.  
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.  
 FT CHAIN 747 809 PROTEIN P7.  
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.  
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.  
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.  
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC -----

FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	476	476	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	STRAND	1224	1226		
FT	TURN	1232	1233		
FT	TURN	1236	1238		
FT	HELIX	1239	1246		
FT	TURN	1247	1248		
FT	STRAND	1251	1255		
FT	HELIX	1258	1271		
FT	TURN	1272	1272		
FT	STRAND	1277	1280		
FT	TURN	1281	1282		
FT	STRAND	1283	1285		
FT	STRAND	1291	1295		
FT	HELIX	1296	1301		
FT	TURN	1302	1303		
FT	STRAND	1312	1316		
FT	TURN	1317	1319		
FT	HELIX	1323	1335		
FT	TURN	1336	1340		
FT	STRAND	1343	1347		
FT	TURN	1352	1353		
FT	TURN	1360	1361		
FT	STRAND	1362	1366		
FT	STRAND	1368	1368		
FT	STRAND	1373	1375		
FT	TURN	1376	1377		
FT	STRAND	1378	1380		
FT	HELIX	1382	1385		
FT	STRAND	1389	1393		
FT	HELIX	1397	1409		
FT	TURN	1410	1411		
FT	STRAND	1414	1417		
FT	TURN	1419	1420		
FT	STRAND	1432	1436		
FT	TURN	1438	1439		
FT	STRAND	1450	1453		
FT	STRAND	1456	1463		
FT	STRAND	1471	1478		
FT	STRAND	1480	1480		
FT	HELIX	1481	1488		
FT	TURN	1489	1490		
FT	STRAND	1497	1501		
FT	STRAND	1507	1507		
FT	STRAND	1511	1511		
FT	HELIX	1514	1527		
FT	HELIX	1532	1544		
FT	STRAND	1550	1550		
FT	HELIX	1555	1564		
FT	HELIX	1570	1578		
FT	TURN	1579	1580		
FT	HELIX	1584	1597		
FT	TURN	1598	1598		
FT	HELIX	1606	1611		
FT	TURN	1614	1618		
FT	STRAND	1622	1623		
FT	STRAND	1627	1627		
FT	STRAND	1635	1636		
FT	HELIX	1640	1652		
SQ	SEQUENCE	3011 AA;	327142 MW;	772CBB29CCD94753	CRC64;

Query Match 86.1%; Score 879.5; DB 1; Length 3011;  
 Best Local Similarity 83.8%; Pred. No. 5.9e-75;

Matches 171; Conservative 9; Mismatches 13; Indels 11; Gaps 2;

QY 3 KKGSVVVG---RIVLNG-----AYAOQTRGEEGCOETSTQGRDNKNQVEGEVQIVST 51  
 DB 1005 RRQETLLGPADQWKGWRLAPITAYAOQTRGLGCTTSLTGDRKNQVEGEVQIVST 1054  
 QY 52 AATQETLATICGVCWTVYHGAGTRTIAASPKGPIOMYTVNDKDLVGVPAPOGSRSLTPCT 111  
 DB 1065 AATQETLATICGVCWTVYHGAGTRTIAASPKGPIOMYTVNDKDLVGVPAPOGSRSLTPCT 1124  
 QY 112 CGSSDLYLVTRHADYIPVRRRGDSRGSLLSPRISYVKGSSGGPILCPAGHAGVIFRAAV 171  
 DB 1125 CGSSDLYLVTRHADYIPVRRRGDSRGSLLSPRISYVKGSSGGPILCPAGHAGVIFRAAV 1184  
 QY 172 CTRGVAKAVDFIPVESLETTMSP 195  
 DB 1185 CTRGVAKAVDFIPVENLETTMSP 1208

RESULT 3

POLG\_HCVTV STANDARD; PRI: 3010 AA.

AC P29846; HCVTV STANDARD; PRI: 3010 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OC Hepacivirus.

OX NCBI\_TaxID=31645;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9220206; PubMed=1314449;

RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;

RT "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.";

RL Virology 188:102-113(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + [RNA](N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M84754; -. NOT\_ANNOTATED\_CDS.

DR PIR; A40244; GNMVTV.

DR PDB; 1N64; 25-FEB-03.

DR PDB; 1NS3; 08-APR-98.

DR MEROPS; S29.001; -.

DR MEROPS; U39.001; -.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR004109; HCV\_NS3.

DR InterPro; IPR000745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RdRP.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.

DR ProDom; PD186062; HCV\_NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).

FT CHAIN 116 191 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/Helicase NS3 (POTENTIAL).

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT CHAIN 347 369 POTENTIAL.

FT TRANSMEM 347 369 POTENTIAL.

FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP\_BIND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC... ) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC... ) (POTENTIAL).

SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 83.0%; Score 847; DB 1; Length 3010;  
 Best Local Similarity 88.2%; Pred. No. 7.1e-72;  
 Matches 157; Conservative 11; Mismatches 10; Indels 0; Gaps 0;





```

DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01003; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR Pfam: PF0186062; HCV_NS1; 1.
DR SMART: SM00487; DEXoc; 1.
DR Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3010 369
FT TRANSMEM 347
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
FT SEQUENCE 3010 AA: 327017 MW: AA993794F46DB185 CRC64;
Query Match 82.1%; Score 838; DB 1; Length 3010;
Best Local Similarity 85.4%; Pred No. 5, le-71;
Matches 152; Conservative 16; Mismatches 10; Indels 0; Gaps 0;
OY 18 AYAOOTRGEQCOEFTSGTRDKNOVEGEVQIVSTAAQTFFLATCINGVGTWVYHCAQGTET 77
Db 1031 AYSOOTRGLGCIITSLGRDKNOVDGEVQLSTATQSLATCVNGVGTWVYHCAQSKTI 1090
OY 78 ASPKGPIVQMTYNDKDLVGVPAPOGSRSLTPTCTCGSSDLVLTTHADVIVPRRGDSRG 137
Db 1091 AGKGPITOMYTNVDLGVWAPPAGKSMPTCTCGSSDLVLTTHADVIVPRRGDSRG 1150
OY 138 SLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRSP 195
Db 1151 SLLSPRISYLGSGGGLLCPGSHVGFIRAVCTRGVAKAVDFIPVESLETTMRSP 1208
RESULT 6
ID POLG_HCVJT STANDARD: PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA] (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC FMBL: D11168; BAA01943.1;
DR PIR: A45573; A45573.
DR DB: IAIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.

```

DE SMART: SM00487; DEXdc; 1.  
 DE Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT CHAIN 3011 369  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 FT SEQUENCE 3010 AA; 326573 MW; 94ALC77435D642BB CRC64;  
 Query Match 82.0%; Score 837; DB 1; Length 3010;  
 Best Local Similarity 86.5%; Pred. No. 6.3e-71;  
 Matches 154; Conservative 13; Mismatches 11; Indels 0; Gaps 0;  
 QY 18 AYAOOTRGEESGCOETSGTRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAGTRTI 77  
 DB 1031 AYAOOTRGLGCVITSLTGRDKNOVEGEVOIVSTAAOTFLATCINGVCWTVYHGAGSKTL 1090  
 QY 78 ASPKGPVITOMYTNVDKLVGNPAFGGSRSLPTCTCGSSDLYLVTRHADVLPVRRRGSRG 137  
 DB 1091 AGPKGPITOMYTNVDQDLVGNHAPPGARSLPTCTCGSSDLYLVTRHADVLPVRRRGSRG 1150  
 QY 138 SLLSPRPVSYLKGSGGPGLLCPAGHVCIFPAAVCTRGVAKAVDFIPVESLETTMSP 195  
 DB 1151 SLLSPRPVSYLKGSGGPGLLCPGSHAVCIFPAAVCTRGVAKAVDFIPVESNETTMSRP 1208  
 RESULT 7  
 ID POLG\_HCVJ6 STANDARD; PRT: 3033 AA.  
 AC P26660;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J6) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9204440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: Comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RT J. Gen. Virol. 72:2697-2704 (1991).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polypeptide, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1',  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: D00944; BAA00792.1; -;  
 DR PIR: JQ1303; JQ1303.  
 DR HSSP: P27958; IHEI.  
 DR MEROPS: S29.001; -;  
 DR MEROPS: 039.001; -;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_Psvir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXdc; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1 1

FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 65.78; Score 670.5; DB 1; Length 3033;  
 Best Local Similarity 62.19; Pred. No. 3.8e-35;  
 Matches 126; Conservative 27; Mismatches 33; Indels 17; Gaps 1;

QY 10 VGRVLNG-----AYAQTRGEGCQETSGTRKDNKQVGEVQIVSTA 52  
 DB 1010 LGREVLGPADCYTSKWSLLAPITAYAQTRGLLGTIVVSMGTGRDKTEOAGEIOVLSTV 1069  
 QY 53 AQTELATRCNGVCTVYHGAGRTIASPGVQIMTYNDVKDLGVGPAPOGSRSLPCTC 112  
 DB 1070 TQSEPLGTTISGVLTVYHGAGNKTLAGSGVPTQMYSSAEGDLGVGPPGPKSLPCTC 1129  
 QY 113 GSSDLVLTNRADVIPVRRKDSRGSLSPRISY:KGSSGGPILCPAGHAVGIFRAAVC 172  
 DB 1130 GAVDLVLTNRADVIPARRKDKRGALLSPRLSLKSSGGPVLCPRHAGVGFRAAVC 1189  
 QY 173 TRGVAKAVDFIPVESLETTMRSP 195  
 DB 1190 SRGVAKSIDFIPVELDIVTRSP 1212

RESULT 8  
 POLG\_HCVJ8  
 ID POLG\_HCVJ8 STANDARD; PRT: 3033 AA.  
 AC P26661;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.  
 OX NCBI\_TaxID=11115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230332; PubMed=1314459;  
 RA Okamoto H., Kural K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.;  
 RT \*Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes.;  
 RL Virology 188:331-341(1992).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D10988; BAA01761.1; -;  
 CC PIR: A40250; GNMVJ8.  
 CC HSSP: P27958; LHEI.  
 CC MEROPS: S29.001; -;  
 CC MEROPS: U39.001; -;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RDRP.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVif.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; Viral\_RDRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.  
 KW INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010



```

FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 1A173E7E3381FDIA CHC64;

Query Match 65.5%; Score 668.5; DB 1; Length 3033;
Best Local Similarity 62.6%; Pred. No. 5.9e-55;
Matches 127; Conservative 25; Mismatches 34; Indels 17; Gaps 1;

QY 10 VGRVLVNG-----AVAQOTCEGECQETSTQGRDKNOVEGHOIVSTA 52
DB 1010 LGREVLGPADGYTSKWKLLAITATYQTQRLGLGAIVVSLTGRDKNOAGQOVLLSV 1069
QY 53 AQTFLATCINGCVTVYHGAGTITASPKGPVIQMTYNDKLVGWPAPOGSRSLTPCTC 112
DB 1070 TOTELGTSGVLVTVYHGAGNCTLAGPKGPVIMTYSAEGDLVGVWPSPGTKSLDPCTC 1129
QY 113 GSSDLXLYLTHADVIPVRRGDSRGLSPRPISYLSKSGGGLLPAGHANGVIFRAVC 172
DB 1130 GAVDLYLTHADVIPVRRGDSRGLSPRPISYLSKSGGGLLPAGHANGVIFRAVC 1189
QY 173 TRGVAKAVDFIPVESLETTMRSP 195
DB 1190 ARGVAKSIDFIPVESLDVATRP 1212

```

## RESULT 9

```

Y136_TREPA
ID Y136_TREPA STANDARD; PRT; 485 AA.
AC O83172;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein tp0136 precursor.
GN TP0136.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=36656576;
RA Fraser C.M., Norris S., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

```

```

RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT *Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete.*;
RL Science 281:375-388(1998).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE TP0136 FAMILY OF LIPOPROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001199; AAC65137.1; ALT_INIT.
CC TIGR: TP0136;
DR Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.
FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 164 178 GLY/SER-RICH.
FT DOMAIN 196 210 GLY/SER-RICH.
FT DOMAIN 253 267 GLY/SER-RICH.
FT DOMAIN 318 327 POLY-SER.
FT DOMAIN 444 447 POLY-SER.
SQ SEQUENCE 485 AA; 48984 MW; C7A4CEEDC7DC5CED CRC64;

Query Match 8.4%; Score 85.5; DB 1; Length 485;
Best Local Similarity 23.6%; Pred. No. 1.3;
Matches 52; Conservative 19; Mismatches 78; Indels 71; Gaps 11;

QY 3 KKGSVIV--GRVNLGAYAAQOTRGECCQETSO----TGRDKNOVEGHOIVSTAQTF 56
DB 53 KAGSKLYATNGRL-----WEKELGTGSGWKVSSSVPTDSK-----KVMSTATDNTF 102
QY 57 LATCI--NGVCWTYVYHGAG---TRTIASPKGPVIQMTYNDKLVG-----WPAPOGSR 105
DB 103 VLACVPGTGVYKHCNVGAGSGSSGTCTASPTETCSOHAT---LVGTSKPEWLVPGGTG 158
QY 106 SLTPCTC-----GSSDLVYLRHADVIP-----VRRGDSRGLSPRPISYLSK--- 149
DB 159 NNGMCGCGGGGGSSSSSCSIHWLVPGTGNNGCGCGGGGGSSSSSSSCSIHIKVEN 218
QY 150 -----GSSGGLPLCPAGHANG 165
DB 219 TDEQFLDMGEGYVYVTKHLYTKNGSSSAGPAQCPCGGGGG 258

```

## RESULT 10

```

HHOA_ARATH
ID HHOA_ARATH STANDARD; PRT; 321 AA.
AC Q9SEL7; O49507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN HHOA OR AT4G18370 OR F28J12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
RT *Identification and characterization of the chloroplast HhoA protease,
RT a homolog to the bacterial periplasmic protease HhoA.*;
RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL

```



[2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Ierlyn N.,  
 RA Harris B., Ansong W., Brandt P., Grivell L., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Bernerstorfer S., Hempel S., Feldpausch M., Lamberth S., van den Daelc H.,  
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Gray-Allen S.,  
 RA Clark L., Daggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Rojandream M.A., Lyne M., Benes V., Reimann S.,  
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehnert I.-H.,  
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi F.,  
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse I.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar R., Jordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Rameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,  
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RI thaliana.";  
 RL Nature 402:769-777(1999).  
 [3]  
 RP SEQUENCE OF 72-82: 96-110; 150-159; 178-211 AND 306-320.  
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,  
 RA Kieselbach T.,  
 RT "The chloroplast lumen from Arabidopsis thaliana.";  
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.  
 CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
 CC gene model prediction. AT4G18370 and AT4G18375 were originally  
 CC fused into a single gene.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF114386; AAF24060.1;  
 CC EMBL: AL021710; CAA16717.1; ALT\_SEQ.  
 CC EMBL: AL161548; CAB78839.1; ALT\_SEQ.

DR MEROPS: S01.279;  
 DR InterPro: IPR001940; Protease2C.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00834; PROTEASES2C.  
 KW Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.  
 FT TRANSIT 1 26  
 FT TRANSIT 27 71  
 FT TRANSIT 72 321  
 FT CHAIN 77 87  
 FT DOMAIN 145 145  
 FT ACT\_SITE 186 186  
 FT ACT\_SITE 264 264  
 FT CONFLICT 40 40 R -> G (IN REF.1)  
 SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;  
 Query Match 8.1%; Score 82.5; DB 1; Length 321;  
 Best Local Similarity 26.2%; Pred. No. 1.6;  
 Matches 39; Conservative 22; Mismatches 49; Indels 39; Gaps 8;  
 QY 73 GRTTASPKGPIQMYNTVNDKLVGWPAPQGSRLTPTCTGSSDLYLVTRHADVIPRRR 132  
 DB 169 GTR--FSKGGKIVGL--DPDNDLAVLKIEGRELNPVVLGTSNDRVGQSCFAI----- 219  
 QY 133 GDSRG-----SLLSPRPISYLK-----GSSGGPLLCAGHAGVIF 167  
 DB 220 GNPYGYENTLTIGVYSGLGREIPSPNGKISIAIOTDADINSNGSGPLDSYGHITGV- 278  
 QY 168 RAACVTR---GVAKAVDF-IPVESLETHM 192  
 DB 279 NTATFTRKSGMSSGVNFAIPDITVVRTV 307  
 RESULT 11  
 AAMP\_HUMAN  
 ID AAMP\_HUMAN STANDARD; PRT: 452 AA.  
 AC Q13685;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Angio-associated migratory cell protein.  
 GN AAMP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=95262124; PubMed=7743515;  
 RA Beckner M.E., Krutzsch H.C., Stracke M.L., Williams S.T.,  
 RA Gallardo J.A., Liotta L.A.;  
 RT "Identification of a new immunoglobulin superfamily protein expressed  
 RI in blood vessels with a heparin-binding consensus sequence.";  
 RL Cancer Res. 55:2140-2149(1995).  
 CC -!- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED  
 CC IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED  
 CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M95627; AAA68889.1;  
 CC PIR: I39383; I39383.  
 CC Genew: HGNC:18; AAMP.  
 CC MIM: 603488; -.

DR GO; GO:0008201; F:heparin binding activity; TAS

DR	InterPro; IPR001680; WD40.	
DR	Pfam; PF00400; WD40; 8.	
DR	SMART; SM00320; WD40; 8.	
DR	PROSITE; PS00678; WD_REPEATS_1; 1.	
DR	PROSITE; PS00082; WD_REPEATS_2; 6.	
DR	PROSITE; PS0294; WD_REPEATS_REGION; 1.	
DR	Repeat; WD repeat.	
FT	DOMAIN	14
FT	FT	18
FT	DOMAIN	71
FT	REPEAT	107
FT	REPEAT	138
FT	REPEAT	150
FT	REPEAT	180
FT	REPEAT	190
FT	REPEAT	220
FT	REPEAT	231
FT	REPEAT	261
FT	REPEAT	306
FT	REPEAT	333
FT	REPEAT	363
FT	REPEAT	374
FT	REPEAT	404
FT	REPEAT	416
FT	REPEAT	446
SQ	SEQUENCE	452 AA; 49015 MW; DA1413D25ER236C0 CRC64;

```

Query Watch      8.0% Score 82; DB 1; Length 452;
Best Local Similarity 25.3% Pred. No. 2.6;
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

QY 66 NTVYHGACTRTIASPKGPIVMYTNVDKDLGVHPAQPGRSL-----TPCTCGSSDLYIV 120
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 197 WNEWH-----PRAPVLLAGT-ADGNTMMKVPNGDCKTFQGNCPATCGR----- 240
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 121 TRHADVIPVR-----GDSRGS-----LLSPRPISYLKGSNG--GPLLCPA----- 160
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 241 -----VLPDGKRAVVGEDCTIRIWLKQGSPIHVLKGTGEGHGLTCAVAAQDGSULT 295
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 161 -----GHAVGTFR-----AAVCTRCVAKAVDFIPVESL 188
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 296 GSDVQCAKLVSATGKYGVGFRTVASQPSLGEGESEESNSVSL 341
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT 12

ID	DEGI_ARATH	STANDARD:	PRT:	437 AA.
AC	O22609;			
AC	O22609; Q9LK95;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).			
GN	DEG1 OR DEGP OR A3G27925 OR K16N12.18.			
OS	Arabidopsis thaliana (Mouse-ear cress);			
OS	Eukaryota; Viridiplantae; Streptophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
OX	[1]			
RP	SEQUENCE FROM N.A.; AND CHARACTERIZATION.			
RP	MEDLINE=98175982; Pubmed=9507020;			
RA	Itzhaki H., Naveh I., Lindahl M., Cook M., Adam Z.;			
RT	Identification and characterization of Degp, a serine protease			
RT	associated with the luminal side of the thylakoid membrane.*;			
RL	J. Biol. Chem. 273:7094-7098(1998).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=cv. Columbia;			
RP	MEDLINE=20363099; Pubmed=10907853;			
RA	Kaneko T., Katoh T., Sato S., Nakamura A., Asanizu E., Tabata S.;			
RT	Structural analysis of Arabidopsis thaliana chromosome 3. II.			
RT	Sequence features of the 4,251,695 bp regions covered by 90 PL. TAC			
RT	and BAC clones.*;			
RL	DNA Res. 7:217-221(2000).			
RL	[3]			
RP	SEQUENCE OF 104-118.			
RP	STRAIN=cv. Columbia;			
RA	Kieselbach T., Bysted M., Schroeder W.P.;			
RT	Submitted (JUL-2000) to the SWISS-PROT data bank.			
CC	-!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.			

CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,  
CC CAN DEGRADE BETA-CASEIN.  
CC -1- ENZYME REGULATION: INHIBITED BY PHENYLMETHYLSULFONYL FLUORIDE AND  
CC O-PHENANTHROLINE.  
CC -1- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID  
CC MEMBRANE.  
CC -1- INDUCTION: By heat shock.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
CC -1- SIMILARITY: Contains 1 PPT/DHR domain.  
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; AF028842; AAC39436.1; -
DR	EMBL; AP000371; BAB02539.1; -
DR	EMBL; AP001302; BAB02539.1; JOINED

DR MEROPS; S01.279; -.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
DR

DR SMARI; SM00226; PDZ; 1.  
DR PROSITE; PS50106; PDZ; 1.  
KW Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.  
FT TRANSIT<sup>1</sup> CHLOROPLAST / POTENTIAL<sup>2</sup>

FT	TRANSIT	?	103	THYLAKOID.
FT	CHAIN	104	437	PROTEASE DO-LIKE 1.
FT	DOMAIN	152	321	SERINE PROTEASE.
FT	DOMAIN	324	421	PDZ.
FT	ACT.SITE	171	171	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT.SITE	201	201	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT.SITE	280	280	CHARGE RELAY SYSTEM (POTENTIAL).
FT	CONFLICT	12	23	HSPPSSQLSNST -> SSTFLHSPSSHL (IN REF.
FT			2)	

FT	CONFLICT	36	V -> I (IN REF. 2).
FT	CONFLICT	36	P -> S (IN REF. 2).
FT	CONFLICT	54	G -> R (IN REF. 2).
FT	CONFLICT	60	G -> D (IN REF. 2).
FT	CONFLICT	64	LL -> HF (IN REF. 2).
FT	CONFLICT	68	L -> V (IN REF. 2).
FT	CONFLICT	355	I -> V (IN REF. 2).
FT	CONFLICT	381	Q -> E (IN REF. 2).
FT	CONFLICT	416	
SO	SPROUCE	437 AA:	1497B1AB33F5FF2A4 CRC64:

Query Match	7.9%	Score 80.5;	DB 1;	Length 437;
Best Local Similarity	25.6%	Pred. No. 3.5;		
Matches	44.	Conservative	18.	Mismatches 55.
				Indels 55.
				Gaps 7.

```
Qy      68 VYHAGTRTIASPKGPVOMY-----TNVDKDLVGW-----PA 100
         | :|: : :|: :|
Pb     150 UNOCSCCCTCTTACGTTATGTCGTCGT DUMI A DOMMTA GPCDGDVGLSIT BTZ 200
```

QY  
101 PQGSRSLTPCTCGSSDIYLIV-----TRHADVIPVRRRGDSRGSLLSPRI 145  
| : | | | | | | | | | | | | | | | |  
:

QY  
146 SYL-----KGSSGPELLCPAGHAVGIFFRAAVCTRGVAKAVDF-IPVESL 180

: : ||||| : | :| | : : |||:::

RESULT 13

IBP1_HUMAN			
ID	IBP1_HUMAN	STANDARD;	PRT; 259 AA.
AC	P08833;		
DT	01-NOV-1988	(Rel. 09, Created)	

DT 01-NOV-1988 (rel. 09, last sequence update)  
DT 28-FEB-2003 (rel. 41, last annotation update)  
DE Insulin-like growth factor binding protein 1 precursor (IGFBP-1)  
DE (IGFBP-1) (IGF-binding protein 1) (placental protein 12) (PP12).  
GN IGFBP1 OR IBP1.  
OS Homo sapiens (Human).  
OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=89052854; PubMed=2461294;  
RA Brinkman A., Groffen C., Kortleve D.J., Geurts A., Drop S.L.S.;  
RT "Isolation and characterization of a cDNA encoding the low molecular  
RT weight insulin-like growth factor binding protein (IGFBP-1).";  
RL EMBO J. 7:2417-2423(1988).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Decidua;  
RX MEDLINE=88240345; PubMed=2454104;  
RA Brewer M.T., Stetler G.L., Squires C.H., Thompson R.C.,  
RA Busby W.H. Jr., Clemmons D.R.;  
RT "Cloning, characterization, and expression of a human insulin-like  
RT growth factor binding protein.";  
RL Biochem. Biophys. Res. Commun. 152:1289-1297(1988).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88335621; PubMed=3419311;  
RA Grundmann U., Nerlich C., Bohn H., Rein T.;  
RT "Cloning of cDNA encoding human placental protein 12 (PP12): binding  
RT protein for IGF I and somatomedin.";  
RL Nucleic Acids Res. 16:8711-8711(1988).  
RN [4]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Decidua;  
RX MEDLINE=88312985; PubMed=2457513;  
RA Julkunen M., Koistinen R., Aalto-Setälä K., Seppälä M., Janne O.A.,  
RA Kontula K.;  
RT "Primary structure of human insulin-like growth factor-binding  
RT protein/placental protein 12 and tissue-specific expression of its  
RT mRNA.";  
RL FEBS Lett. 236:295-302(1988).  
RN [5]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=88334540; PubMed=2458522;  
RA Lee Y.-L., Hintz R.L., James P.M., Lee P.D.K., Shively J.E.,  
RA Powell D.R.;  
RT "Insulin-like growth factor (IGF) binding protein complementary  
RT deoxyribonucleic acid from human HEP G2 hepatoma cells: predicted  
RT protein sequence suggests an IGF binding domain different from those  
RT of the IGF-I and IGF-II receptors.";  
RL Mol. Endocrinol. 2:404-411(1988).  
RN [6]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=8930502; PubMed=2474129;  
RA Cubbage M.L., Suwanichkul A., Powell D.R.;  
RT "Structure of the human chromosomal gene for the 25 kilodalton  
RT insulin-like growth factor binding protein.";  
RL Mol. Endocrinol. 3:846-851(1989).  
RN [7]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=89087480; PubMed=2849945;  
RA Brinkman A., Groffen C.A., Kortleve D.J., Drop S.L.S.;  
RT "Organization of the gene encoding the insulin-like growth factor  
RT binding protein IGF-1.";  
RL Biochem. Biophys. Res. Commun. 157:898-907(1988).  
RN [8]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=92217971; PubMed=1373120;  
RA Ehrenborg E., Larsson C., Stern I., Janson M., Powell D.R.,

RA Luthman H.;  
RT "Contiguous localization of the genes encoding human insulin-like  
RT growth factor binding proteins 1 (IGBP1) and 3 (IGBP3) on chromosome  
RT 7.";  
RL Genomics 12:497-502(1992).  
RN [9]  
RN SEQUENCE OF 141-259 FROM N.A., AND SEQUENCE OF 26-259.  
RC TISSUE=Amniotic fluid;  
RX MEDLINE=89170723; PubMed=2466665;  
RA Luthman H., Soederling-Barros J., Persson B., Engberg C., Stern I.,  
RA Lake M., Franzen S.A., Israelsson M., Raden B., Lindgren B.,  
RA Hjelmqvist L., Enerbaeck S., Carlsson P., Bjursell G., Pova G.,  
RA Hall K., Joernvall H.;  
RT "Human insulin-like growth-factor-binding protein. Low-molecular-mass  
RT form: protein sequence and cDNA cloning.";  
RL Eur. J. Biochem. 180:259-265(1989).  
RN [10]  
RN SEQUENCE OF 26-53.  
RX MEDLINE=89008261; PubMed=2971653;  
RA Busby W.H. Jr., Klapper D.G., Clemmons D.R.;  
RT "Purification of a 31,000-dalton insulin-like growth factor binding  
RT protein from human amniotic fluid. Isolation of two forms with  
RT different biologic actions.";  
RL J. Biol. Chem. 263:14203-14210(1988).  
RN [11]  
RN MUTAGENESIS.  
RX MEDLINE=92070504; PubMed=1959616;  
RA Brinkman A., Kortlirve D.J., Schuller A.G.P., Zwarthoff E.C.,  
RA Drop S.L.S.;  
RT "Increase of beta-actin mRNA upon hypotonic perfusion of perfused rat  
RT liver.";  
RL FEBS Lett. 292:264-268(1991).  
RN [12]  
RN PHOSPHORYLATION SITES, AND PARTIAL SEQUENCE.  
RX MEDLINE=93123224; PubMed=7678248;  
RA Jones J.I., Busby W.H. Jr., Wright G., Smith C.E., Kinack N.M.,  
RA Clemmons D.R.;  
RT "Identification of the sites of phosphorylation in insulin-like  
RT growth factor binding protein-1. Regulation of its affinity by  
RT phosphorylation of serine 101.";  
RL J. Biol. Chem. 268:1125-1131(1993).  
RN [13]  
RN DISULFIDE BONDS.  
RX MEDLINE=99262603; PubMed=10329650;  
RA Neumann G.M., Bach L.A.;  
RT "The N-terminal disulfide linkages of human insulin-like growth  
RT factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as  
RT determined by mass spectrometry.";  
RL J. Biol. Chem. 274:14587-14594(1999).  
CC -!- FUNCTION: IGF-binding proteins prolong the half-life of the IGFs  
CC and have been shown to either inhibit or stimulate the growth  
CC promoting effects of the IGFs on cell culture. They alter the  
CC interaction of IGFs with their cell surface receptors.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- PTM: Phosphorylated; probably by casein kinase II. Alters the  
CC affinity of the protein for IGFs.  
CC -!- MISCELLANEOUS: Binds equally well IGF-I and IGF-II.  
CC -!- SIMILARITY: Contains 1 IGFBP domain.  
CC -!- SIMILARITY: Contains 1 thyroglobulin type-1 domain.  
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts  
CC in positions 55 and 71.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Y00856; CAA68770.1; -;  
DR EMBL; M20841; AAA52540.1; ALT\_FRAME.  
DR EMBL; X12385; CAA30942.1; -;

EMBL: X13405; CAA31771.1; -  
 DR EMBL: M31145; AAA52542.1; -  
 DR EMBL: M59316; AAA52783.1; -  
 DR EMBL: M23595; AAA52785.1; -  
 DR EMBL: M23592; AAA52785.1; JOINED.  
 DR EMBL: M23593; AAA52785.1; JOINED.  
 DR EMBL: M23594; AAA52785.1; JOINED.  
 DR EMBL: M23587; AAA52784.1; -  
 DR EMBL: X15002; CAA33110.1; -  
 DR EMBL: A31867; IQHUL.  
 DR HSP: P24593; LBOE.  
 DR TRANSFAC: T00400; -  
 DR Slena-2DPAGE; P08933; -  
 DR Genew; HGNC:5469; IGFBP1.  
 DR MIM; 146730; -  
 DR GO; GO:0005615; Extracellular space; TAS.  
 DR GO; GO:0005520; F-insulin-like growth factor binding activity; TAS.  
 DR GO; GO:0007165; P-signal transduction; TAS.  
 DR InterPro; IPR000867; Ins\_lgro\_fac.pr.  
 DR Pfam; PF00219; IGFBP; 1.  
 DR Pfam; PF00086; thyroglobulin\_1; 1.  
 DR SMART; SM00121; IB; 1.  
 DR SMART; SM00211; TY; 1.  
 DR PROSITE; PS00222; IGF\_BINDING; 1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; 1.  
 KW Growth factor binding; Signal; Phosphorylation; Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 259 INSULIN-LIKE GROWTH FACTOR BINDING  
 FT DOMAIN 202 251 PROTEIN 1.  
 FT SITE 246 248 THYROGLOBULIN TYPE I.  
 FT DISULFID 71 84 CELL ATTACHMENT SITE.  
 FT DISULFID 78 104  
 FT DISULFID 176 206  
 FT MOD\_RES 126 126 PHOSPHORYLATION (MAJOR).  
 FT MOD\_RES 144 144 PHOSPHORYLATION.  
 FT MOD\_RES 194 194 PHOSPHORYLATION.  
 FT VARIANT 183 183 V -> I (IN dBSNP:1065782).  
 FT VARIANT 253 253 I -> M (IN dBSNP:4619).  
 FT VARIANT 253 253 /FTID-VAR\_003821.  
 FT CONFLICT 213 213 H -> Q (IN REF. 2).  
 FT SEQUENCE 259 AA; 27903 MW; 8AA75AF7DC966012 CRC64;

Query Match 7.7%; Score 79; DB 1; Length 259;

Best Local Similarity 28.3%; Pred. No. 2.6;

Matches 26; Conservative 12; Mismatches 26; Indels 28; Gaps 4;

QY 12 RIVLNCAYAQOTRGEK-----GCOFIS--QTGRDKNOVEGFVQIVSTAAQTFLATCIN 62

DB 182 RVVSLAKAGETSSEISKEYLPNCNKNGFHSROCTSMDEA----- 225

QY 63 GVCWTYVHGAGTRTIAASPK---GPVTOMYTNV 91

DB 226 GLWCVCYPPNGKRIPGSPFIRGDPNCQIYENV 257

RESULT 14

PAAD\_PSEAE

ID PAAD\_PSEAE STANDARD; PRT; 209 AA.

AC Q9HX08;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable aromatic acid decarboxylase (EC 4.1.1.-).

GN PA4019.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Hickey C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Stover C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goulet L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Harbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT \*Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.\*;  
 RL Nature 406:959-964(2000).  
 CC -!- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /  
 CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

EMBL: AE004818; AAG07406.1; -

PIR: H83144; H83144.

InterPro: IPR003382; Flavoprotein.

Pfam: PF02441; Flavoprotein; 1.

KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.

SQ SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;

Query Match 7.7%; Score 78.5; DB 1; Length 209;

Best Local Similarity 26.5%; Pred. No. 2.3;

Matches 50; Conservative 16; Mismatches 56; Indels 67; Gaps 11;

QY 41 QVEGEVO-IVSTAAQTFLATCINGVCWTVYHGAGTRTIAASPKGP----- 83

DB 29 QEEREVHFLSKAAQLVNAI-----ETDVALPAKPAQAPLTYCGAAAG 74

QY 84 VIQMYTNVDKLVGPAPQGSRLTP-----CTCGSSDL-----YLVTRHADVIPV 129

DB 75 QIRVFGQND-----WMAPPASGSSAPNAWICPCSTGLTSAVATGACNNLIERAADVALK 129

QY 130 RRRGDSRGSLLSPR--PIS-----YLGSSGGFLLCPAGHVCIFRAAVCTRGVAKVD 181

DB 130 ER----RPLVLPREAPFSSIHLENMLKLSNIGAVILPA--APGFYHQ---POSVEDLVD 180

QY 182 FIPVESLET 190

DB 181 FVVARILNT 189

RESULT 15

ENV\_FLVGL

ID ENV\_FLVGL STANDARD; PRT; 642 AA.

AC P08359;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein

DE GP70; Spike protein P15E].

GN ENV.

OS Feline leukemia virus (strain A/Glasgow-1).

OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.

OX NCBI\_TaxID=11769;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86200439; PubMed=3009890;

RA Stewart M.A., Warnock M., Wheeler A., Wilkie N., Mullins J.I.,

RA Onions D.E., Nell J.C.;

RT \*Nucleotide sequences of a feline leukemia virus subgroup A envelope

RT gene and long terminal repeat and evidence for the recombinational

RT origin of subgroup B viruses.\*;

RL J. Virol. 58:825-834(1986).



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.2105 seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKKGSVYIGRIYNGAYA.....VAKAYDFIPVESLETTMRSP 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-ARCHEA:\*
- 2: SP-BACTERIA:\*
- 3: SP-FUNGI:\*
- 4: SP-HUMAN:\*
- 5: SP-INVERTEBRATE:\*
- 6: SP-MAMMAL:\*
- 7: SP-PLANT:\*
- 8: SP-ORGANELLE:\*
- 9: SP-PHAGE:\*
- 10: SP-PLANT:\*
- 11: SP-RODENT:\*
- 12: SP-VIRUS:\*
- 13: SP-VERTEBRATE:\*
- 14: SP-UNCLASSIFIED:\*
- 15: SP-VIRUS:\*
- 16: SP-BACTERIAP:\*
- 17: SP-ARCHEAP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908.5	89.0	4040	12 Q91FH8	Q91fh8 mucosal dis
2	894.5	87.6	2436	12 Q81756	Q81756 hepatitis c
3	894.5	87.6	3011	12 Q91FE5	Q91fe5 hepatitis c
4	893.5	87.5	3011	12 Q03463	Q03463 hepatitis c
5	891	87.3	181	12 Q91RR8	Q91rr8 hepatitis c
6	891	87.3	181	12 Q91RT5	Q91rt5 hepatitis c
7	889.5	87.1	3011	12 Q36579	Q36579 hepatitis c
8	889	87.1	181	12 Q91RR5	Q91rr5 hepatitis c
9	888	87.0	181	12 Q91RR3	Q91rr3 hepatitis c
10	888	87.0	181	12 Q91RS1	Q91rs1 hepatitis c
11	888	87.0	181	12 Q91RQ8	Q91rq8 hepatitis c
12	888	87.0	181	12 Q91RT1	Q91rt1 hepatitis c
13	886	86.8	181	12 Q91RR6	Q91rr6 hepatitis c
14	886	86.8	181	12 Q91RS9	Q91rs9 hepatitis c
15	885.5	86.7	3011	12 Q91RS8	Q91rs8 hepatitis c
16	885	86.7	181	12 Q91RR2	Q91rr2 hepatitis c

17	885	86.7	181	12 Q91RS3	Q91rs3 hepatitis c
18	884	86.6	181	12 Q91RT4	Q91rt4 hepatitis c
19	884	86.6	181	12 Q91RS8	Q91rs8 hepatitis c
20	884	86.6	181	12 Q91RT3	Q91rt3 hepatitis c
21	884	86.6	181	12 Q91RS5	Q91rs5 hepatitis c
22	884	86.6	181	12 Q91RS7	Q91rs7 hepatitis c
23	884	86.6	181	12 Q91RT0	Q91rt0 hepatitis c
24	882.5	86.4	3011	12 Q9DT6	Q9dt6 hepatitis c
25	882.5	86.4	3011	12 Q36608	Q36608 hepatitis c
26	882.5	86.4	3011	12 Q9PW5	Q9pwx5 hepatitis c
27	882.5	86.4	3015	12 Q9PW9	Q9pwx9 hepatitis c
28	882	86.4	181	12 Q91RS4	Q91rs4 hepatitis c
29	881	86.3	181	12 Q91RT6	Q91rt6 hepatitis c
30	880	86.2	181	12 Q91RT9	Q91rt9 hepatitis c
31	879	86.1	181	12 Q91RR4	Q91rr4 hepatitis c
32	879	86.1	181	12 Q91RR9	Q91rr9 hepatitis c
33	879	86.1	181	12 Q91RR0	Q91rr0 hepatitis c
34	877	85.9	181	12 Q91RR7	Q91rr7 hepatitis c
35	876.5	85.8	3011	12 Q36609	Q36609 hepatitis c
36	876	85.8	181	12 Q91RT2	Q91rt2 hepatitis c
37	876	85.8	181	12 Q91RR1	Q91rr1 hepatitis c
38	876	85.8	181	12 Q91RQ9	Q91rq9 hepatitis c
39	876	85.8	181	12 Q91RS2	Q91rs2 hepatitis c
40	874	85.6	181	12 Q91RS6	Q91rs6 hepatitis c
41	873.5	85.6	3011	12 Q36610	Q36610 hepatitis c
42	873	85.5	181	12 Q91RT7	Q91rt7 hepatitis c
43	871	85.3	181	12 Q91RS0	Q91rs0 hepatitis c
44	871	85.3	181	12 Q91RT8	Q91rt8 hepatitis c
45	860.5	84.3	3011	12 Q81754	Q81754 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q91FH8	PRELIMINARY:	PRT: 4040 AA.
AC	Q91FH8:		
DT	01-OCT-2000 (TReMBLrel. 15, Created)		
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TReMBLrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OX	Pestivirus.		
OX	NCBI_TaxID=111099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20323484; PubMed-10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassiliev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	"Generation and Characterization of a Hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.";		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF268278; AAF92566.1; ..		
DR	HSSP: P26663; 1XP.		
DR	MEROPS: S31.001; ..		
DR	InterPro: IPR000280; CDvir_endptsep80.		
DR	InterPro: IPR001410; HCV_NS3.		
DR	InterPro: IPR004109; HCV_RdRP.		
DR	InterPro: IPR002166; HCV_RdRP.		
DR	InterPro: IPR001650; Helicase_C.		
DR	InterPro: IPR001005; Myb_DNA_Binding.		
DR	InterPro: IPR001568; RNase_T2.		
DR	InterPro: IPR007095; RNA_pol_DS_PS.		
DR	InterPro: IPR007094; RNA_pol_PSVir.		
DR	Pfam: PF02907; HCV_NS3; 1.		
DR	Pfam: PF00271; Helicase_C; 1.		
DR	Pfam: PF00998; Viral_RdRP; 1.		

```

DR PRINTS: PR00729; CDVENDOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS00507; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
DR PROSITE; PS00531; RNASE_T2_2; 1.
DR PROSITE; PS00531; RNASE_T2_2; 1.
DR PROSITE; PS00531; RNASE_T2_2; 1.
KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA; 453073 MW; ADE87791D053B9DC CRC64;

Query Match      89.08; Score 908.5; DB 12; Length 4040;
Best Local Similarity 90.88; Pred. No. 9e-84;
Matches 177; Conservative 6; Mismatches 7; Indels 5; Gaps 1;

OY 5 GSVVIVGRVLNG-----AYAQOTRGEEGCQETSTQTRGKNGVEGEVQIVSTAAQIFLAT 59
DB 10 GSVVIVGRVLNGSGSSTACAAQOTRGLGCKRTITSLTRGKNGVEGEVQIVSTATQIFLAT 69
OY 60 CINGVCWTVYHGAGRTIASPKGPVQMTYNDKDLVGPAPQGSRLTPTCTCGSSDLYL 119
DB 70 CINGVCWTVYHGAGRTIASPKGPVQMTYNDKDLVGPAPQGSRLTPTCTCGSSDLYL 129
OY 120 VTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKA 179
DB 130 VTRHANVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKA 189
OY 180 VDFIPVESLETTMRS 194
DB 190 VDFIPVENLETTMRS 204

RESULT 2
O81756
ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX Choo Q.-L., Richman K., Han J.;
RT *The nucleotide sequence of the Hepatitis C viral genome.*;
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -
DR HSSP; P27958; IALV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR Pfam; PF018062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00507; RDRP_POSITIVE; 1.

DR PROSITE; PS05021; RDRP_VIRAL; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
DR Hydrolase; Nonstructural protein; Polyprotein;
DR RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 2436 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match      87.68; Score 894.5; DB 12; Length 2436;
Best Local Similarity 85.88; Pred. No. 1.3e-82;
Matches 175; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

OY 3 KGSVVIVG---RIVLNG-----AYAQOTRGEEGCQETSTQTRGKNGVEGEVQIVST 51
DB 555 PRGRILLGPADGMVSGWRLAPITAYAQOTRGLGCKRTITSLTRGKNGVEGEVQIVST 614
OY 52 AAQFLATCINGVCWTVYHGAGRTIASPKGPVQMTYNDKDLVGPAPQGSRLTPTCT 111
DB 615 AAQFLATCINGVCWTVYHGAGRTIASPKGPVQMTYNDKDLVGPAPQGSRLTPTCT 674
OY 112 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 171
DB 675 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 734
OY 172 CIRGVAKAVDFIPVESLETTMRS 195
DB 735 CIRGVAKAVDFIPVENLETTMRS 758

RESULT 3
O91FE5
ID Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21262212; PubMed-11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT *Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT sequence.*;
RL J. Gen. Virol. 82:1291-1297(2001).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; IALV.
DR InterPro; IPR000345; Cyt_c_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.

```





```
QY 172 CTRGVAKAVDFIPVESLETMRSP 195
Db 1185 CTRGVAKAVDFIPVESLETMRSP 1208

RESULT 5
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC STRAIN=Pt.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF36235; AAK54560.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19130 MW; 85091869299B7C35 CRC64;

Query Match 87.3%; Score 891; DB 12; Length 181;
Best Local Similarity 96.6%; Pred. No. 1.le-83;
Matches 171; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAAQTGEGCGQETSQTRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRTI 77
Db 5 AYAAQTGGLGCIITSLTGRDKNQVEGEVQIVSTAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPGKGVQIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSRG 137
Db 65 ASPGKGVQIOMYTNVDKDLVGPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSRG 124
QY 138 SLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVESLETMR 194
Db 125 SLLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETMR 181

RESULT 7
Q36579 PRELIMINARY; PRT; 3011 AA.
AC Q36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RC STRAIN=H77;
RA MEDLINE-97373636; PubMed-9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_core; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01338; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
```

```
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS050521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transforase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0E809G63C1B9 CRC64;

Query Match      87.1%; Score 889.5; DB 12; Length 3011;
Best Local Similarity 84.8%; Pred. No. 5.6e-82;
Matches 173; Conservative 9; Mismatches 11; Indels 11; Gaps 2;

QY 3 KKGSVIVG---RVING-----AYAOQTRGEGCOETSGTRDKNOVEGEVQIVST 51
DB 1005 RGQELLGPRDGMYSKGRLLAPITAYAOQTRGLLGCIITSLTGROKNOVEGEVQIVST 1064
QY 52 AQOTFLATCINGCVTVYHGAGTRTIASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCT 111
DB 1065 ATOTFLATCINGCVTVYHGAGTRTIASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCT 1124
QY 112 CGSSDLYLVTRHADYIPVRRRGRSGSLSPRISYLGSSGGPLLCPCPAGHAGVIFRAAV 171
DB 1125 CGSSDLYLVTRHADYIPVRRRGRSGSLSPRISYLGSSGGPLLCPCPAGHAGVIFRAAV 1184
QY 172 CTRGVAKAVDFIPVESLETTMRS 195
DB 1185 CTRGVAKAVDFIPVENLETTMRS 1208

RESULT 8
Q91RR5 PRELIMINARY; PRT; 181 AA.
ID Q91RR5;
AC Q91RR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP STRAIN=Pt.30;
RC Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369238; AAK54563.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F21C0A72 CRC64;

Query Match      87.1%; Score 889; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 1.8e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAOQTRGEGCOETSGTRDKNOVEGEVQIVSTAAQTFLATCINGCVTVYHGAGTRTI 77
DB 5 AYAOQTRGLLGCIITSLTGROKNOVEGEVQIVSTAAQTFLATCINGCVTVYHGAGTRTI 64
QY 78 ASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCTCGSSDLYLVTRHADYIPVRRRGRSG 137
DB 65 ASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCTCGSSDLYLVTRHADYIPVRRRGRSG 124
QY 138 SLLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 194
DB 125 SLLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 9
Q91RR3 PRELIMINARY; PRT; 181 AA.
ID Q91RR3;
AC Q91RR3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP STRAIN=Pt.4B;
RC Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369240; AAK54565.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19115 MW; 5D85F88AD7AC1A11 CRC64;

Query Match      87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AYAOQTRGEGCOETSGTRDKNOVEGEVQIVSTAAQTFLATCINGCVTVYHGAGTRTI 77
DB 5 AYAOQTRGLLGCIITSLTGROKNOVEGEVQIVSTAAQTFLATCINGCVTVYHGAGTRTI 64
QY 78 ASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCTCGSSDLYLVTRHADYIPVRRRGRSG 137
DB 65 ASPKGPVIOYTNVDKDLVGPAPGQSRSLTPCTCGSSDLYLVTRHADYIPVRRRGRSG 124
QY 138 SLLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 194
DB 125 SLLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 10
Q91RS1 PRELIMINARY; PRT; 181 AA.
ID Q91RS1;
AC Q91RS1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP STRAIN=Pt.K;
RC Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369232; AAK54557.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match      87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
```

```
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMR 181

RESULT 11
Q91RQ8 ID Q91RQ8 PRELIMINARY; PRT: 181 AA.
AC Q91RQ8:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN-Pt.52;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369245; AAK54570.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19144 MW; C0C91F1E2EEB0B32 CRC64;

Query Match 87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMR 181

RESULT 12
Q91RT1 ID Q91RT1 PRELIMINARY; PRT: 181 AA.
AC Q91RT1:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-Pt.161;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369222; AAK54547.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 87.0%; Score 888; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 2.3e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMR 181

RESULT 13
Q91RR6 ID Q91RR6 PRELIMINARY; PRT: 181 AA.
AC Q91RR6:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.3T;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369237; AAK54562.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19101 MW; 614ADA8B0F33CCAF CRC64;

Query Match 86.8%; Score 886; DB 12; Length 181;
Best Local Similarity 95.5%; Pred. No. 3.7e-83;
Matches 169; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 18 AYAAQTRGEGCCQETSOTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 77
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 AYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRTI 64
QY 78 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 ASPKGPVIQMTYNTVDKDLVGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPIVRRGDSRG 124
QY 138 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMR 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 SLLSPRPISYLGSSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMR 181
```

```
RESULT 14
Q91RS9      PRELIMINARY;      PRT: 181 AA.
AC Q91RS9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Pt.174;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF369224; AAK54549.1;
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DB0635 CRC64;

Query Match      86.8%; Score 886; DB 12; Length 181;
Best Local Similarity 96.0%; Pred. No. 3.7e-83;
Matches 170; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 AVAAQOTRGECCOETSGTGRDNQVGEVQIVSTAAQIFLATCINGCVWTVYHGAGTRTI 77
DB 5 AVAAQOTRGILGCIITSLTRGDKDQVGEVQIVSTAAQIFLATCINGCVWTVYHGAGTRTI 64

QY 78 ASPKGPVIOYMYNVDKDLVGMWPAQGSRSRLTCTCGSSDLXLVTRHADVPVRRRGDSRG 137
DB 65 ASPKGPVIOYMYNVDKDLVGMWPAQGSRSRLTCTCGSSDLXLVTRHADVPVRRRGDSRG 124

QY 138 SLISPRPISYLGSSGGPLICPAGHAGVIFRAAVCFIRGAKAVDFIPVESLETTMRS 194
DB 125 SLISPRPISYLGSSGGPLICPAGHAGVIFRAAVCFIRGAKAVDFIPVENLETTMRS 181

RESULT 15
Q9ELS8      PRELIMINARY;      PRT: 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF290978; AAG02099.1;
DR HSP; P27958; IHE1.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
```

```
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match      86.7%; Score 885.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.4e-81;
Matches 172; Conservative 10; Mismatches 11; Indels 11; Gaps 2;

QY 3 KKGSVTVWG---RVLNG-----AVAQOTRGECCOETSGTGRDNQVGEVQIVST 51
DB 1005 RGQEILLGPADGWVSGWRLAPITAYAAQOTRGLGCIITSLTRGDNQVGEVQIVST 1064

QY 52 AAQIFLATCINGCVWTVYHGAGTRTIASPKGPVIOYMYNVDKDLVGMWPAQGSRSRLT 111
DB 1065 ATQIFLATCINGCVWTVYHGAGTRTIASPKGPVIOYMYNVDKDLVGMWPAQGSRSRLT 1124

QY 112 CGSSDLXLVTRHADVPVRRRGDSRGSLSPRPISYLGSSGGPLICPAGHAGVIFRAAV 171
DB 1125 CGSSDLXLVTRHADVPVRRRGDSRGSLSPRPISYLGSSGGPLICPAGHAGVIFRAAV 1184

QY 172 CTRGVAKAVDFIPVESLETTMRSP 195
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

Search completed: August 30, 2003, 19:18:18
Job time : 38.2105 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: August 30, 2003, 15:18:33 ; Search time 2534.57 seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-12  
Perfect score: 1021  
Sequence: 1 MKKGSVIVIGRIVLINGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Searched: 2888711 scqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO\_spool/US09965594/runat\_29082003.151919.28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=pct -THR.WAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN\_1\_1.14686\_@runat\_29082003.151919.28310 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6  
-YGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hgt: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pt: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_on: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_to: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pin: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_hgtgo\_hum: \*  
40: em\_hgtgo\_mus: \*  
41: em\_hgtgo\_other: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	928.5	90.9	12734	6	AR179057 Sequence
2	908.5	89.0	12734	14	AF268278 Pestiviru
3	894.5	87.6	5360	6	AR118686 Sequence
4	894.5	87.6	5360	6	I06434 Sequence 48
5	894.5	87.6	5360	6	I09328 Sequence 8
6	894.5	87.6	6785	6	AR118692 Sequence
7	894.5	87.6	6785	6	I06440 Sequence 54
8	894.5	87.6	6785	6	I09329 Sequence 10
9	894.5	87.6	7310	6	AR118696 Sequence
10	894.5	87.6	7310	6	I09331 Sequence 15
11	894.5	87.6	7310	14	HPCPOLYP M32084 Hepatitis C
12	894.5	87.6	8316	6	AR118703 Sequence
13	894.5	87.6	8987	6	AR118728 Sequence
14	894.5	87.6	9185	6	AR118722 Sequence
15	894.5	87.6	9185	6	AR118723 Sequence
16	894.5	87.6	9185	6	BD091382 HCV cult1
17	894.5	87.6	9185	6	I08294 Sequence 1
18	894.5	87.6	9379	6	AR166930 Sequence
19	894.5	87.6	9379	6	AR301300 Sequence
20	894.5	87.6	9401	6	AR176483 Sequence
21	894.5	87.6	9401	6	BD080334 Hepatitis C
22	894.5	87.6	9401	6	E66593 Hepatitis C
23	894.5	87.6	9401	6	I71894 Sequence 9
24	894.5	87.6	9401	6	I81885 Sequence 9
25	894.5	87.6	9401	14	HPCPOLYPRE M62321 Hepatitis C
26	894.5	87.6	9609	12	AF387805 Synthetic
27	894.5	87.6	9609	12	AF387808 Synthetic
28	894.5	87.6	9618	14	AF271632 Hepatitis C
29	894.5	87.6	9646	12	AF387806 Synthetic
30	894.5	87.6	9693	12	AF387807 Synthetic
31	894	87.6	2058	6	AX395309 Sequence
32	894	87.6	2058	6	AX454818 Sequence
33	893.5	87.5	9502	6	E08263 Sequence
34	893.5	87.5	9502	6	E08264 CDNA of Hep
35	893.5	87.5	9502	14	HPCHCJ1 D10749 Hepatitis C
36	892	87.4	1932	6	AR127809 Sequence
37	892	87.4	1932	6	BD081910 Hepatitis
38	892	87.4	8157	6	AR127810 Sequence
39	892	87.4	8157	6	BD081911 Hepatitis
40	891.5	87.3	1998	6	AR145264 Sequence
41	891.5	87.3	9424	14	AF511948 Hepatitis
42	891	87.3	543	14	AF369218 Hepatitis
43	891	87.3	543	14	AF369235 Hepatitis
44	890	87.2	2061	6	AX441176 Sequence
45	890	87.2	2061	6	AX467113 Sequence

ALIGNMENTS

RESULT 1

AR179057	AR179057	Sequence 1	12734 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AF268278.1	from patent US 6326137.				
DEFINITION	AF268278.1	GI:20220612				
ACCESSION	AF268278.1	GI:20220612				
VERSION	AF268278.1	GI:20220612				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 12734)					
AUTHORS	Hong, Z., Lai, V.C.H. and Lau, J.Y.N.					
TITLE	Hepatitis C virus, protease-dependent chimeric pestivirus					
JOURNAL	Patent: US 6326137-A 1 04-DEC-2001;					
FEATURES	Location/Qualifiers					
source	1..12734					
BASE COUNT	4032 a 2604 c 3295 g 2803 t					
ORIGIN						
Alignment Scores:						
Pred. No.:	3,06e-66	Length:	12734			
Score:	928.50	Matches:	180			
Percent Similarity:	94.87%	Conservative:	5			
Best Local Similarity:	92.31%	Mismatches:	5			
Query Match:	90.94%	Indels:	5			
DB:	6	Gaps:	1			
US-09-965-594-12 (1-195) x AR:79057 (1-12734)						
Qy	5	GlySerValValIleValGlyArgGluValLeuAsnGly	-----AlaTyr 19			
Db	413	GGTAGTGTGTTATGTTGTAGATGTTTATCTGGTAGTGTAGTATCAGCGGTAC	472			
Qy	20	AlaGlnGlnThrArgGlyGluGluGlyCysGlnClnuThrSerGlnThrGlyArgAspLys	39			
Db	473	GCCACGACGAGAGGCGCTCTAGGGTGTAGATCACCAGTCTGACTGCGCGGGACAAA	532			
Qy	40	AsnGlnValGluGlyValGlnIleValSerThrAlaAlaGlnGlnThrPheLeuAlaThr	59			
Db	533	AACCAAGTGAGGAGTGGTCCAGATCGTGTCAACTGCTACCAACACCTTCCTGGCAACG	592			
Qy	60	CysIleAsnGlyValCysTsrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer	79			
Db	593	TGCATCAATGGGTATGTGTGACTGTCTACACCGGGCGGACGAGGACCATCGCATCA	652			
Qy	80	ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro	99			
Db	653	CCCAAGGTCCTGTCTATCCAGATATATACCAATGTGACCAAGACCTTGTGGCTGGCCC	712			
Qy	100	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	119			
Db	713	GCCTCTCAAGGTTCCTGCTATGACACCTGACCTGCGGCTCGGACCTTTACCTG	772			
Qy	120	ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	139			
Db	773	GTTACGAGGACCGCGACGCTATCCCGTGGCGGAGGTATACGAGGGTAGCCTG	832			
Qy	140	LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro	159			
Db	833	CTTTCGCGCGCGCCATTCTCTACCTAAAAGGCTCTCTCGGGGGTCCGCTGTGTGCCCC	892			
Qy	160	AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla	179			
Db	893	CGGGAGACCGCGTGGCGCTATTACAGGGCGCGGTGTGCACCGGTGAGTGGCCAAAGCG	952			
Qy	180	ValAspPheIleProValGluSerLeuGluThrThrMetArgSer	194			
Db	953	GTGACATTTATCCTGTGGAGAACCTTAGACAAACCATGAGATCC	997			
RESULT 2						
LOCUS	AF268278	12734 bp	RNA	linear	VRL 12-JUL-2000	
DEFINITION	Pestivirus type 1, complete genome.					

ACCESSION	AF268278					
VERSION	AF268278.1	GI:9049956				
KEYWORDS	Pestivirus type 1					
SOURCE	Pestivirus type 1					
ORGANISM	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Pestivirus.					
REFERENCE	1 (bases 1 to 12734)					
AUTHORS	Lai, V.C., Zhong, W., Skelton, A., Ingravallo, P., Vassiliev, V., Donis, R.O., Hong, Z. and Lau, J.Y.					
TITLE	Generation and characterization of a hepatitis C virus NS3 protease-dependent bovine viral diarrhea virus					
JOURNAL	J. Virol. 74 (14), 6339-6347 (2000)					
MEDLINE	20323484					
PUBMED	10864644					
REFERENCE	2 (bases 1 to 12734)					
AUTHORS	Lai, V.C.H. and Hong, Z.					
TITLE	Direct Submission					
JOURNAL	Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA					
FEATURES	Location/Qualifiers					
source	1..12734					
	/organism="Pestivirus type 1"					
	/mol_type="genomic RNA"					
	/db_xref="taxon:11099"					
	1..385					
5'UTR	386..12508					
CDS	/codon_start=1					
	/product="polyprotein"					
	/protein_id="AAF82566.1"					
	/db_xref="GI:9049957"					
	/translation="MELNTNGSSGVIVGRIVLSSGSGITACAAQTRGLGCKITSL					
	TGRDNQVEGVQIVSTATQFLATCINGVCVTVYHGAGTRTIAASPKGPIQMTYND					
	QDLGWAPAQSSRLTPTCTGSSDLYLVTRHANVIVRRGDSRGLSPRIPLYNG					
	SSGGLLCAGHAGVGLFRAACTGCAKADFPDPVENLETTTSGSDAGTDDVYCCSM					
	SYSDTKEGATKKTKQDRLEGRKMKVPKSEKDSKTKPPDATIVVEGVGYVRKK					
	GKTKSNTQDGLYHNKPKQSRKKLEKALLAWAIIAIVLQVINGENITQWLDNG					
	TEGIRAFKRGVNSRHGIMPEKICTGYPSHLATDIEKTHGMDASLNTTCR					
	LORHEMFKGNCWNYNIEPWLVMNRTQANLTGEGOPPRECAVTCYDRASDLNVYTA					
	RDSPTLTGCKGRNFSFAGILMRGPNCFEAASDLVFKHERISMFOQDTLYLVGL					
	TNSLEGARQGTAKTLTWLGKILGKLENKSKTFWGAASPYCDVDRKIGYIWT					
	KNTFACLPKNTKIVGPKFDTNADGKILHEMGHLSVLLLSLVLSLDRAPETASV					
	MYLLISFIPSHVDVMDCKTQNLNLTVELITADVPGSVNLCWVLRINMPYET					
	TVLAFEVSQVVKLVLRALDRLTRINNAATTAFLCLVKIVRGQMVGILNLLIT					
	VQGHDCCKPEFYAIAKDERIGQLGAEGLTTWKEYSPGMKLEMTYIACEDGKLM					
	YLQCTRETRYLAHLRALPTSVYKXLFDRKQEDVYEMNDNFEPGLCPDCAPIV					
	RQKPTTLTRNGFAQVQCPIGTGTVTSCTSFNMDTLATTVYTRSRSPFPHROQIT					
	QKNLEDHNCILGNWTCVPGDOLLYKGGSTESCKWGYOFKESGLPHYPIGCKL					
	ENETCYRLVSTSCNREGVAIVPGCTLCKKIGKTTVOVIAQMDTKLGPMPCRPYETSS					
	EGPVKTAQTFNYTKTLNKKYEPEDRSYFQQYMLAGEYQYWFLEVDLTHHRDYAESI					
	LYVVALLGGRVLMWLLVTYIVLSOKALGIQYSGSEVVMGNLTHNNIEVYTFLL					
	LYVLLRESEYKWKWLLLYHILVHPISKVIVILLMIGDVVKADSGGQBYLQKIDCFT					
	TVYLVIGLIIAARDPTIVPLTMAALRVLTETHQPGVDIAVAVMTILMLSVYTD					
	VYRYKKWQCILSLSVGLIRSLYVLAGRIEMPEVTIPNWRPLTILLYLISLTIVTR					
	WRVDYAGLLQCPILLVLTLMADFLILILPTVELVKLVYLTATVTDTERSWLGG					
	IDYTRVDSIYDDESGEYVLPSPQKAGNFIILLPLIKATLISCVSKWQLIYMSY					
	LTLDPMYTHRKVIEISGNTIISRIAAALIELMWSEESKGLKAKFILLSGLRN					
	LIHKYMTNRVTASVYGEVYMPKIMTIKASTLSKSRHICITCTCEGRENKGGTC					
	PKCGHGXPTCCMSLADFEERYKRIETRENGFCMGSCQCKRRRFPDREKSHAR					
	YCAECNRLHPAEGDGFWAEESMLGLITFALMDKGYDITFWAGCQVRGSPDTHR					
	PCHISFGSMPFRQYNGVQYARGQLFNLNPLVATKYKMLVGNLGEENLEHL					
	GWILKGPVCKAITIEHKCHINLDKTAFFGIMPRGTPRAPVFPFTSLKVRKGLE					
	TGMATYHOGGTSVDVHTAGKDLLVDSMGRTRVYQSNRNLTDTEYGVKTDSCPD					
	GARCVLNPFAVNIISGKGVYHMQTGTEFTCTVASTGPAFFDLNKLKMGSLPIFE					
	ASSGVRVGVKVNKESKPTKMSGIQTVSKNTADLTVMKKITSMRKGDFKOITLA					
	TCAGTTELPKAVIEIENHKKRVLLILRAAESVYQYMRKLRHPSIFNLRIDMKE					
	GDWATGTTASGYFCQPPQPKRAAMVEYSITFLDETHCATPEQALIIKIHRES					
	IRVAMATPAGSVTTGQKHPIEEFIAPEVNKMGDSQFQDIAGLKIPIVDMRGNM					
	LVFVPTNRNMAVEAKKAKGNSGYTSGEDPANLRVYTSQSPYIVATVAIESGYT					
	LPDQTVTDGLCEKRVVSKIPFIVTGLKMAVTVGEQARRGVGRVPGPYR					
	SQETATGSKDYHYDLQAQRYGIEDGINVTKSFREMYDWSLYREDSLLITQLELNN					
	LLISDLPANVNIINARDHPPIQIOLAVNSYEVQVPLFPKIRNGEYTDVTSFNL					
	ARKUGEDVPVIYIATEDEDELVDLLDMDPDPGNOQVETGKALKKQVDTGLSAAENLL					

3'UTR  
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others  
ORIGIN  
VALGYYGQALSKRRVPMITDIYTIEDORLEDTHLOYAPNAIKTDCGTETELKELAS  
GDVEKIMGAI5DYAAGLEFVKQAEKIKTAPLEKENAEAKGVQVRFIDSLKENKE  
IIRYCLMGHTFALYKSTAAIRLGHETAFATLVIKW:AFGGSVSDBHVKCAADVLYVYV  
LNKPSFPDQSDTQEGRRFVSLFISALATYKTYKWNHNSKVDEPALAYLPVATSV  
MKMFTPTRLSSVILSTIYKTYLSIRKGSXSDG:ILCTCISAAFEILSONPVSGVSM  
LGVGAIAHNAIESEQKRTLMMKVFYKFNFLDOAA:DELKYENPEKILMALFEAVQII  
YIIDLIVHLHGVIYKWEAKELSERAGRNLFILIMPEAFELLMDSQGRKILNSG  
YIIDLIVHLHGVIYKWEAKELSERAGRNLFILIMPEAFELLMDSQGRKILNSG  
EMKAFKGGKLTVEESGPFICRNRRPGPVNYRVYKYDDNLRREIKPVAKLEGQVE  
HYKGVTAIDYKSGKMLATDKWEVEHGVITRLAKRYTGVGNGAYLVGDPNHRALV  
PRDCATITKNTVOFLKMKKCAFTYDITISNLTJLJELVHNNLFEKEIPTATVTWML  
AYTFVNEVDGILKPVLSERVIPDPVDINLOPEVOQVTSVGIIGRETLMTIGVTP  
VLEKVEPDASONSQVIGLDEGNYPGGIOQTHTLTETIEHNRDARPEFTMILGNSIS  
NRAKTARINITYNDPREIRDLMAAGRMVVALRDVDPSELSEMYDFKTLIDREAL  
ALSJGQPKVKTVEAVENLIEQKQDEIPNMFASDDPFLVLEKALNDKYVLVGDVGE  
VKDQALGATDTRLIKEVGSRTYAMKLSWFQASNKOMSLTPLELLRCPPTAT  
KSNKGHASAVOLAGQWNEPIGCGVHLCTIPARVYIHPYEAYLKLDFTFEERKKPR  
VKDVIREHNNKILKIRFOGNLNTKMLNPKGLSEQLDREGRNRNIHQIGTINS  
AGIRLEKLIPIVRAOTDTTIFHEAIRDKIDKSENRONPELHKKLLEIFHTIAOPTLKHT  
YGEVTEQLEAGINRKGAGFELEKKNIGEIDSEKHLVEQLVRLDKAGRIKXYETAI  
PKNKRQVSDQWQADLVVEKRPVIOYPPAKTRIAITKVHYNVVKQOPVPIGVEGK  
TPLFNIPOKVRKDSNPVAVSFDTKAMDVTQSKDLOIIGFIOKYKKKWHKFI  
DTJDMTEFVPIITADGEVYIRNGORSGCOPDSAGNSMLNVLTMVAFCSCTGVYK  
SENRVARIHVGCDGFLITKGLGLKFKANKGWOILHEAGKPKQITEGKKMVAIFED  
IEFCSHTPVPVMSDSSHAGRDITAVILSKMATRLDSSGGERGTATYKAVAFSFL  
MYSNNPLVRRICLLVLSQOPETDPSKHATYYKGDPIGAYKDVIGRNLSELRKTFEK  
LANLNLSTICTITWKTSKRIIQDCAVIGKEGNLWVADRI:ISSKTGLYIPDKGF  
TIQGHYEQLOLRTEPNVMVGCTERYKLGPIVNL:LLRRLKILLMTAVGSS\*  
12509, 12734  
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,35e-64 Length: 12734  
Score: 908.50 Matches: 177  
Percent Similarity: 93.8% Conservative: 6  
Best Local Similarity: 90.77% Mismatches: 7  
Query Match: 88.98% Indels: 5  
DB: 14 Gaps: 1  
US-09-965-594-12 (1-195) x AF268278 (1-12734)  
QY 5 GlySerValValIleValGlyArgIleValLeuAsnGly-----AlaTyr 19  
DB 413 GGTAGTGTGTATTGTGGTAGAATGTTTATCTGCTAGTGTAGTATCATCGCGTGC 472  
QY 20 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 39  
DB 473 GCCCAGCAGCAGAGAGGCTCTCAGGGTGTAGATCATCAGCTGCTGACTGGCGGGACAAA 532  
QY 40 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 59  
DB 533 AACCAAGTGGAGGTCAGGTCAGATCGTGCTCACTGCTACCCAAACCTTCTGCGCAACG 592  
QY 60 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 79  
DB 593 TGCATCAATGGGGTATGCTGCACTCTCTACACCGGGCGGGAACGAGCATCCATCA 652  
QY 80 ProLysGlyProValIleGlnMetThrThrAsnValAspLysAspLeuValGlyTrpPro 99  
DB 653 CCCAAGGTCCTGTCATCAGATGATACCAATGTGGACAGACCTTGTGGGCTGGGCC 712  
QY 100 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 119  
DB 713 GCTCTCAAGTTCCTGCTCATGTGACACCTGCACCTGCGGCTCTCTCGACCTTTACCTG 772  
QY 120 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 139  
DB 773 GTTACGAGCAGCCCAACGTCATTCCTGCGCGCGCAGAGTGATAGCAGGGGTAGCCTG 832  
QY 140 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 159  
DB 833 CTTTCGCCCGGCCCATTTCTACCTAAAGAGCTCTCTGCGGGTCCGCTGTGTGCGCC 892

QY 160 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 179  
DB 893 CGCGGACACGCCGTGGCGCTATTTCAGGCGCGGCTGTGCACCGGTGAGTGGCCAGCG 952  
QY 180 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194  
DB 953 GTGACTTTATCCCTGTGGAGNACCTAGACACACCAGGATCC 997  
RESULT 3  
ARL18686  
LOCUS ARL18686 5360 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 53 from patent US 6150087.  
ACCESSION ARL18686  
VERSION ARL18686.1 GI:14100596  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Chien,D.Y.  
TITLE NABV diagnostics and vaccines  
JOURNAL Patent: US 6150087-A 53 21-NOV-2000;  
FEATURES Location/Qualifiers  
source 1..5360  
/organism="unknown"  
BASE COUNT 1060 a 1623 c 1532 g 1145 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.42e-64 Length: 5360  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
DB: 6 Gaps: 2  
US-09-965-594-12 (1-195) x ARL18686 (1-5360)  
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
DB 867 CGCAGGGCGGGAGATACTGCTCGGGCCACCGCATGGATGGTCTCCAGGGGGTGAGG 926  
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31  
DB 927 TTGCTGGCGCCCATCACGGCGTACGCCAGCAGACAGAGGCCCTCCTAGGGTGCTAATC 986  
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
DB 987 ACCAGCCTAATGCGCGGACAAAACCAAGTGGAGGGTGAGGTCAGATTGTGCTCAACT 1046  
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
DB 1047 GCTGCCCAACCTTCTCTGGCAACGTGCATCAATGGGGTGTGCTGACTGTCTACACGGG 1106  
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
DB 1107 GCGGAAGCAGGACCATCGCTACCCNAGGGTCTCTGCTCATCAGATGTATACCAATGTA 1166  
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
DB 1167 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCCCTGCACT 1226  
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
DB 1227 TCGGGCTCTCGGACCTTTACCTGGTCACGAGGACCGCATGTCATTCGCGTCCGCGCG 1286  
QY 132 ArgGlyAspSerArgGlySerLeuSerProArgProIleSerTyrLeuLysGlySer 151  
DB 1287 CGGGGTGATACAGGGGCGCTCTGCTGCGCGCGGCCCATTTCTACTTGAAGAGCTCC 1346  
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171

Db 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGACAGCGGTGGGCATATTTAGGGCCCGGGTG 1406  
QY 172 CysThrArgGlyValAlaIysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
Db 1407 TGCACCCGTGGAGTGGCTAAGGGGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466  
QY 192 MetArgSerPro 195  
Db 1467 ATGAGGTCCCG 1478  
RESULT 4  
LOCUS 106434 5360 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 48 from Patent EP 0318216.  
ACCESSION 106434  
VERSION 106434.1 GI:590311  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton,M., Choo,Q.-I., and Kuo,G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: EP 0318216-A1 48 31-MAY-1989;  
FEATURES Location/Qualifiers  
source  
1..5360  
/organism="unknown"  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.42e-64 Length: 5360  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
Gaps: 2  
US-09-965-594-12 (1-195) x 106434 (1-5360)  
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
Db 867 CGCAGGGCCGGAGATACCTGCTCGGCCACCGCATGGAATGCTCTCCAAGGGTGGAGG 926  
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlnGlu 31  
Db 927 TTGCTGGCGCCCATCACGGGTACGCCACGACAGAGGGGCTCTAGGGTGCATAATC 986  
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
Db 987 ACCACCTTAACCTTCCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1046  
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
Db 1047 GCTGCCCAACCTTCCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1106  
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
Db 1107 GCCGGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1166  
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCAATTGACACCCCTGCAC 1226  
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
Db 1227 TGGCGCTCCTCGGACCTTTACCTGGTCAGAGGACGCCGATGTCATTCCTCGCGCGG 1286  
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
Db 1287 CGGGGTGATAGAGGGGACGCTGCTGCCCGCGGCCCATTTCTACTTTCAAAGGCTCC 1346  
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171

Db 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGACAGCGGTGGGCATATTTAGGGCCCGGGTG 1406  
QY 172 CysThrArgGlyValAlaIysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
Db 1407 TGCACCCGTGGAGTGGCTAAGGGGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466  
QY 192 MetArgSerPro 195  
Db 1467 ATGAGGTCCCG 1478  
RESULT 5  
LOCUS 109328 5360 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 8 from Patent WO 8904669.  
ACCESSION 109328  
VERSION 109328.1 GI:587963  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton,M., Choo,Q.-K., and Kuo,G.  
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;  
FEATURES Location/Qualifiers  
source  
1..5360  
/organism="unknown"  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.42e-64 Length: 5360  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
Gaps: 2  
US-09-965-594-12 (1-195) x 109328 (1-5360)  
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
Db 867 CGCAGGGCCGGAGATACCTGCTCGGCCACCGCATGGAATGCTCTCCAAGGGTGGAGG 926  
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlnGlu 31  
Db 927 TTGCTGGCGCCCATCACGGGTACGCCACGACAGAGGGGCTCTAGGGTGCATAATC 986  
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
Db 987 ACCACCTTAACCTTCCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1046  
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
Db 1047 GCTGCCCAACCTTCCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGCTACCAAGGG 1106  
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
Db 1107 GCCGGAACGAGGACCATCGGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1166  
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCAATTGACACCCCTGCAC 1226  
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
Db 1227 TGGCGCTCCTCGGACCTTTACCTGGTCAGAGGACGCCGATGTCATTCCTCGCGCGG 1286  
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
Db 1287 CGGGGTGATAGAGGGGACGCTGCTGCCCGCGGCCCATTTCTACTTTCAAAGGCTCC 1346  
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171



```
|||||
Db 1347 TCGGGGGTCCGCTGTTGGCCCCGGGACGCGGTGGGCATATTATAGGGCCGCGGTG 1406
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 1407 TGCACCGGTGGAGTGGCTAGAGCGGTGACATTATCCCTGTGGAGAACCTAGAGACAACC 1466
Qy 192 MetArgSerPro 195
Db 1467 ATCAGGTCCCGC 1478

RESULT 6
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
    source
        1. .6785
        /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9,59e-64 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x AR118692 (1-6785)
Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1140 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGCATGGTATGCTCCAGGGGTGGAGG 1199
Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1200 TTGCTGGCGGCCATACGCGGTACGCCAGCAGACAAGGGGCTCCTAGGGTGCATAATC 1259
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 1260 ACCAGCCTAACTCGCGCGGCAAAAACCAAGTGGAGGTGAGTCCAGATTGTCTCACT 1319
Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTyrThrValTyrHisGly 71
Db 1320 GCTGCCCAAACTTCTCGCAAGCTGCATCAATGGGGTGTCTGGACTGTCTACACGGG 1379
Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1380 GCCGAACGAGGACCATCGCGTCACCAAGGGTCTGTATCCAGATGATATACCAATGTA 1439
Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1440 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 1499
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTTCAGGAGCAGCGCATGTCTCCCGTGGCGCGG 1559
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1560 CGGGGTGATAGCAGGGGCAAGCTGCTGTCGCCCCGCGCCCATTTCTTACTTGAAGGCTCC 1619
```

```
Qy 152 SerClyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaIaVal 171
Db 1620 TCGGGGGTCCGCTGTTGTCCCGCGGACGCGGTGGGCATATTATAGGGCCGCGGTG 1679
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 191
Db 1680 TGCACCGGTGGAGTGGCTAGAGCGGTGACATTATCCCTGTGGAGAACCTAGAGACAACC 1739
Qy 192 MetArgSerPro 195
Db 1740 ATCAGGTCCCGC 1751

RESULT 7
LOCUS I06440 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 54 from Patent EP 0318216.
ACCESSION I06440
VERSION I06440.1 GI:590312
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,O.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 54 31-MAY-1989;
FEATURES
    source
        1. .6785
        /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9,59e-64 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x I06440 (1-6785)
Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1140 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGCATGGTATGCTCCAGGGGTGGAGG 1199
Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1200 TTGCTGGCGGCCATACGCGGTACGCCAGCAGACAAGGGGCTCCTAGGGTGCATAATC 1259
Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 1260 ACCAGCCTAACTCGCGCGGCAAAAACCAAGTGGAGGTGAGTCCAGATTGTCTCACT 1319
Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTyrThrValTyrHisGly 71
Db 1320 GCTGCCCAAACTTCTCGCAAGCTGCATCAATGGGGTGTCTGGACTGTCTACACGGG 1379
Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1380 GCCGAACGAGGACCATCGCGTCACCAAGGGTCTGTCTCCAGATGATATACCAATGTA 1439
Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1440 GACCAAGACCTTGTGGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAC 1499
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTTCAGGAGCAGCGCATGTCTCCCGTGGCGCGG 1559
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1560 CGGGGTGATAGCAGGGGCAAGCTGCTGTCGCCCCGCGCCCATTTCTTACTTGAAGGCTCC 1619
```

```

Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 1620 TCGGGGGTCCGCTGTGTGCCCCGGCGGACCGCGTGGCATATTATTAGGGCCGGGTG 1679

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGlySerLeuGluThrThr 191
Db 1680 TGCACCCGTGGAGTGGCTAAGCGGGTGGACITTTATCCCTGTGGAGAACCTAGAGACAACC 1739

Qy 192 MetArgSerPro 195
Db 1740 ATGAGGTCCCGG 1751

RESULT 8
LOCUS 109329
DEFINITION Sequence 10 from Patent WO 8904669.
ACCESSION 109329
VERSION 109329.1 GI:587964
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,Q.-K. and Kuo,G.
JOURNAL Patent: WO 8904669-A 10 01-JUN-1989;
FEATURES
Location/Qualifiers
source 1..6785
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 9,59e-04 Length: 6785
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x 109329 (1-6785)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1140 CGCAGGGCGGGAGATACGTCTGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 1199

Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1200 TTGCTGGCGCCCATCACGGCGGTACGCCAGCAGACAGGGGCCCTCTTAGGGTGCATATC 1259

Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 1260 ACCAGCCTAAGTGGCGGGACAAAACCAAGTGGAGGGTGGAGTCCAGATTGTGCAACT 1319

Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1320 GCTGCCCAAACTTCTTGGCAACGTGCATCAATGGGTGTGCTGGAGTGTCTACACCGG 1379

Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1380 GCCGGAACGAGGACCATCGGTACCCACCAAGGGTCTGTCTCCAGATGTATACCAATGTA 1439

Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1440 GACCAAGACCTTGTGGGTGGCGGCTCCGCAAGGTAGCGGCTCATTCACACCCCTGCAC 1499

Qy 112 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 1500 TCGCGCTCTCGGACCTTACCTGGTCACGAGGACGCCGA?GTCA?TTCGGTCCGCGCGG 1559

Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
Db 1560 CGGGGTGATAGCAGGGGAGCGCTGCTGTGCGCCCGGCCCATTTCTCTACTTGAAGGGTCC 1619

```

```

Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 1620 TCGGGGGTCCGCTGTGTGCCCCGGCGGACCGCGTGGCATATTATTAGGGCCGGGTG 1679

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGlySerLeuGluThrThr 191
Db 1680 TGCACCCGTGGAGTGGCTAAGCGGGTGGACITTTATCCCTGTGGAGAACCTAGAGACAACC 1739

Qy 192 MetArgSerPro 195
Db 1740 ATGAGGTCCCGG 1751

RESULT 9
LOCUS 109329
DEFINITION Sequence 74 from patent US 6150087.
ACCESSION 109329
VERSION 109329.1 GI:14100606
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 7310)
AUTHORS Chien,D.Y.
JOURNAL NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 74 21-NOV-2000;
FEATURES
Location/Qualifiers
source 1..7310
BASE COUNT 1495 a 2220 c 2056 g 1539 t
ORIGIN

Alignment Scores:
Pred. No.: 1.04e-63 Length: 7310
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x AR118696 (1-7310)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 1665 CGCAGGGCGGGAGATACGTCTGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 1724

Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31
Db 1725 TTGCTGGCGCCCATCACGGCGGTACGCCAGCAGACAGGGGCCCTCTTAGGGTGCATATC 1784

Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51
Db 1785 ACCAGCCTAAGTGGCGGGACAAAACCAAGTGGAGGGTGGAGTCCAGATTGTGCAACT 1844

Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71
Db 1845 GTCGCCCAAACTTCTTGGCAACGTGCATCAATGGGTGTGCTGGAGTGTCTACACCGG 1904

Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 1905 GCCGGAACGAGGACCATCGGTACCCACCAAGGGTCTGTCTCCAGATGTATACCAATGTA 1964

Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 1965 GACCAAGACCTTGTGGGTGGCGGCTCCGCAAGGTAGCGGCTCATTCACACCCCTGCAC 2024

Qy 112 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 131
Db 2025 TGGCGCTCTCGGACCTTACCTGGTCACGAGGACGCCGA?GTCA?TTCGGTCCGCGCGG 2084

Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151

```

Db 2085 CGGGGTGATAGCAGGGGCGAGCTGCTGTGCGCCCGCGGCCCATTTCTTACTTGAAGAGCTCC 2144

Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
|||||  
Db 2145 TCGGGGGTCCGCTGTGTCGCCCGGGGCGACGCCGTGGGCATATTTAGGCGCGGGT 2204

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
|||||  
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAACC 2264

Qy 192 MetArgSerPro 195  
|||||  
Db 2265 ATGAGTCCCG 2276

RESULT 10

LOCUS I09331 7310 bp DNA linear PAT 02-DEC-1994

DEFINITION Sequence 15 from Patent WO 8904669.

ACCESSION I09331

VERSION I09331.1 GI:587966

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 7310)  
Houghton, M., Choo, Q.-K. and Kuo, G.  
Patent: WO 8904669-A 15 01-JUN-1989;

FEATURES  
Location/Qualifiers  
source 1..7310  
/organism="unknown"  
BASE COUNT 1495 a 2218 c 2058 g 1539 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1 04c-63 Length: 7310  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x I09331 (1-7310)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
:::||||| :|||

Db 1665 CGCAGGGCGCGGAGATAGTCTCGCGCGCGGATGGAATGCTCCAAAGGGT\*CGAGG 1724

Qy 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31  
|||||

Db 1725 TTGCTGGCGCCCATCAGCGGTACGCCCGCAGCAGACAGGGGGCTCTCTAGGGTGCATAATC 1784

Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnValGlnIleValSerThr 51  
|||||

Db 1785 ACCAGCTTACTGCGCGCGGCAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACT 1844

Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
|||||

Db 1845 GCTGCCCAACCTTCTCGCAACGTGCATCAATGAGGTGCTGTGGAGTCTGTACACCGC 1904

Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetThrThrAsnVal 91  
|||||

Db 1905 GCCCGAAGCAGGACCATCGCGTCACCAAGGTCCTGTCTATCCAGATGATACCAATGTA 1964

Qy 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
|||||

Db 1965 GACCAAGACCTTGTGGGTGGCCCGCTCCGCAAGTAGCGGCTCATGTACACCTGCAT 2024

Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
|||||

Db 2025 TCGCGCTCTCGACCTTTTACCTGGTCCAGGACGACGCCGATGTCATTCCTCGCGCGG 2084

Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
|||||

Db 2085 CGGGGTGATAGCAGGGGCGAGCTGCTGTGCGCCCGCGGCCCATTTCTTACTTGAAGAGCTCC 2144

Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
|||||  
Db 2145 TCGGGGGTCCGCTGTGTCGCCCGGGGCGACGCCGTGGGCATATTTAGGCGCGGGT 2204

Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
|||||  
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAACC 2264

Qy 192 MetArgSerPro 195  
|||||  
Db 2265 ATGAGTCCCG 2276

RESULT 11

LOCUS HPCPOLYP 7310 bp ss-RNA linear VRL 02-AUG-1993

DEFINITION Hepatitis C virus polyprotein gene, partial cds.

ACCESSION M32084

VERSION M32084.1 GI:329875

KEYWORDS polyprotein.

SOURCE Hepatitis C virus

ORGANISM Hepatitis C virus

REFERENCE 1 (bases 1 to 7310)  
Choo, Q.-L., Richman, K. and Han, J.  
The nucleotide sequence of the Hepatitis C viral genome  
Unpublished (1990)  
Original source text: Hepatitis C virus, cDNA to viral RNA, Clones  
K9-1 through 15e, isolated from chimpanzee (individual 910) blood  
plasma.  
Draft entry and printed sequence for [1] kindly submitted by  
M.Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street,  
Emeryville CA 94608.

FEATURES  
Location/Qualifiers  
source 1..7310  
/organism="Hepatitis C virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:11103"  
<1..>7310  
/note="polyprotein"  
/codon\_start=3  
/protein\_id="AAA45677.1"  
/db\_xref="GI:329876"  
/translation="GCERL\*ASCRPLTDFQGGWPISYANGSGPDQRPYCMWHPKPC  
GIVPAKSVGGVYCTPSPVVGTTDRSGAPTSWGENDTDFVFNNTRPPLGNWFC  
TWNSTGCTKVCAGAPCVIGGAGNNTLHCPDCEPKHPDATYSCRGSGPWITPRCLVD  
YPRLWHPCTINVTIFKIRMYGVGVEHLEAFCNMTGRSCDLEDKRSFELSLIT  
ITQOVLPCEFTTLPALSTGLIHQNLVDVQLYGVGSSIASHAIKWEYVWLLFLLL  
ADARVCSCLMMLLSQAEALENIVILNAASLAGTHGLVSLVFFCFAMTLKGVVP  
CAVTFYCMWPLLLILLALPORAYALDTEVAASCGVVLVGLMALTLSPPYKRYISWC  
LWMLQYFLTRVEAQLHVMIPPLNVGRGRDAVILLCAVHPTLVFDITKLLAVFGLP  
ILOALLKVLGVFVQVGLRFLCALARKMIGHYVQVLIKLGLTGYVYNHLTPLRD  
WAHNGRLDVAVEPVFVSQMETKLITWGADTAACGDIINGLPSVARSERILLGAD  
GMVSGRWLLAPITAYAOQTGGLGGLTISLTGRDKNOVEVEVIVSTAQTFLATCI  
NGVCTVTHGAGTHTIASPGPVIOMYTNVDQDLGWPAQPSGSLTPTCTGSSDLVL  
TRHADVTPRRRGDSRGLSPRPISYLKSGSGGLPCPAGHAGVTFRAACTGRTVA  
KAVDFIPVENLETTMRSPVTDNSPPVPOVAHLHAPTGSKTKYPAAYAAOG  
KYVLVLPVSAATLGGNAYMSKAGIDPNIRTGVTITTGSPITYSTYGLFADGGS  
GGADITDCEDSTDATSILGTVLDQAEIAGRLVLAATATPPSVVYVPHNISE  
VALSTGEIPYKRAIPLVITKGRHLIFHSKKCDELAKLVALGINAVINAYTRGLD  
VSVTPSGDVVAVATDAIMTGTGDFSDVDCNTCTOTVDFSLDPTFTTETITLPD  
ASVTRQRTGRGKPGIYREVPAPGRSGFSDSVLCECVGACWYELTPTATYCAR  
LRAYNTPLPCVODHLEFEGVFTGLTHDAHFLSQTQKSGENLPYLVAQATVCAR  
AQAPPSSDQMWKCLIRLKPTLHGPTPLLYBLGAVONEITLTHPTKYINTCMADLE  
VYTSWLVGGVLAALAAAYCLSTGCIVGVGRVLSGRPAIIPDEVLYRFEDEKEGS  
QHLPIEQQMMLAEQFKKALGLLOQTASROAEVIAVQTNWQKLETFMAKWNWFIIS  
GIQLAGLSLTPGPAISMAFTAAVTSPITSTQTLFNLIGWVAAQALAAQCAATA  
FVGAGLCAAGIAGSVGLKVLIDILAGYAGAGVAGALVAFKIMSGVSPSTEDVNLPLAI  
LSPGALVGVVCAAILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPESDAARV  
TALLSLIVTQLRLRHQWISSBCTPCSGSNLRDINDWICEVLSDFKTLWKLKMLPQ  
LPGIPFVSCQRGYKGVWRVDGIMHTRCHGAEITGVKNKGMTIRVIGPRTCKRNWFS

PINAYTTGCTTLPAPNTYFALWRVSAEYVEIROVGFHYVTGMTDNLKCPCQVPS  
PEFFTELDGVRHLRAPPCLLREEVFRVGLHEYPVGSQLPCEPEPDVAVLTSMLT  
DPSHTAEAGRRRLARGSPSVASSASOLSAPSLKATCTANHDSFDAELIGANLWR  
QSMGNIITVESENKVVILDSFDPLVAFEDEREISVPAEILBKSPRFAQALPVWARP  
YNPLVETWKKDYEPPVHVGGCLPLPPKSPPPVPRKRTVVLTSTLSTALAEATR  
SFGSSSTGIDGNTLTYTSEPAPSCPDSDAEYSMPPLEGEPDGLDSDSWSTY  
SSEANEDYVCCSMYSMTGALVTPCAAEOKLPINALNSLLRHINLIVSTTSAC  
QOKKVTFOVLQVLDHYQDLVEKAAAKVKNLLSVEEACSLTPHSAKSKFYG  
ANDRCHARKMAYTHINSVKKLLEDNVTPIDTTINAKNEVFCVQPEKGRKRPALIV  
PDGLVPCVKMALYDVVTKLPLAVMCSVSYFOYSPQORVERFLVQAWKSKKTPMGFSYD  
TRCFDSTVTSERDITEEALYQCCDDIDPOARVAIKSLTERLYVGGPLTNSRGNCGYRR  
CRASGVLTSCSNLTYCYIKARAARAGLQDCTMLVCGDDLLVICESAGVEDAASL  
RAFTAEATRYSPAPGPPPEYDLELITSCSNVSAHHDGAKRVYITRTDPTTLAR  
AAWETARHFPVNSWLGNIITWFAPTLWARMILMTFFSVLIARDQLEQALDCBIYAGY  
SIEPLDLPIIOLR\*

BASE COUNT 1495 a 2218 c 2058 g 1539 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,04e-63 Length: 7310  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
DB: 14 Gaps: 2

US-09-965-594-12 (1-195) x HPGPOLYP (1-7310)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
Db 1665 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGATGGAAATGGTCTCCAAGGGGTGGAGG 1724  
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyGlnGlu 31  
Db 1725 TTGCTGGCGCCCATCAGCGGTACGCCACACAAAGGGGCTCTAGGTGCATAATC 1784  
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
Db 1785 ACCAGGCTAACTGGCGGGCAAAACCAAGTGGAGGTGAGGTCCAGATTTGTCAACT 1844  
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
Db 1845 GCTGCCCAAACTCTCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGTCTACCAAGGG 1904  
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
Db 1905 GCCGGAACAGGACCATCGCGTCACCAAGGGTCTGTCTATCCAGATGATACCAATGTA 1964  
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
Db 1965 GAGCAAGACCTTGTGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAT 2024  
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
Db 2025 TGGGCTCTCGGACCTTTACCTGGTCACGAGCAGCGCGATGTCATTCCTGGCGCGG 2084  
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
Db 2085 CGGGGTATAGCAGGGCAGCTGCTGTGCGCGCGGCCCATTTCTACTTTGAAAGGCTCC 2144  
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
Db 2145 TCGGGGGTCCGCTGTGTGCGCGCGGGCAGCGCGTGGGCATATTATGAGGGCGCGGTG 2204  
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
Db 2205 TGCACCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACNACC 2264  
QY 192 MetArgSerPro 195  
Db 2265 ATGAGGTCCCGG 2276

RESULT 12

AR118703  
LOCUS AR118703 8316 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 88 from patent US 6150087.  
ACCESSION AR118703  
VERSION AR118703.1 GI:14100613  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 8316)  
AUTHORS Chien,D.Y.  
TITLE NANV diagnostics and vaccines  
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;  
FEATURES Location/Qualifiers  
1..8316  
/organism="unknown"

BASE COUNT 1671 a 2529 c 2345 g 1771 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.2e-63 Length: 8316  
Score: 894.50 Matches: 175  
Percent Similarity: 89.71% Conservative: 8  
Best Local Similarity: 85.78% Mismatches: 10  
Query Match: 87.61% Indels: 11  
DB: 6 Gaps: 2

US-09-965-594-12 (1-195) x AR118703 (1-8316)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
Db 2671 CGCAGGGCGGGAGATACTGCTCGGGCCAGCGATGGAAATGGTCTCCAAGGGGTGGAGG 2730  
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyGlnGlu 31  
Db 2731 TTGCTGGCGCCCATCAGCGGTACGCCACAGACAGAGGGGCTCTAGGTGCATAATC 2790  
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
Db 2791 ACCAGGCTAACTGGCGGGCAGCAAAACCAAGTGGAGGTGAGGTCCAGATTTGTCAACT 2850  
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
Db 2851 GCTGCCCAAACTCTCTGGCAAGTGCATCAATGGGGTGTGCTGGACTGTCTACCAAGGG 2910  
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
Db 2911 GCCGGAACAGGACCATCGCGTCACCAAGGGTCTGTCTATCCAGATGATACCAATGTA 2970  
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
Db 2971 GACCAAGACCTTGTGGCTGGCGGCTCCGCAAGGTAGCGGCTCATTTGACACCTGCAT 3030  
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
Db 3031 TGGGCTCTCGGACCTTTACCTGGTCACGAGCAGCGCGATGTCATTCCTGGCGCGG 3090  
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
Db 3091 CGGGGTATAGCAGGGCAGCTGCTGTGCGCGCGGCCCATTTCTACTTTGAAAGGCTCC 3150  
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
Db 3151 TCGGGGGTCCGCTGTGTGCGCGCGGGCAGCGCGTGGGCATATTATGAGGGCGCGGTG 3210  
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
Db 3211 TGCACCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCTTAGAGACNACC 3270  
QY 192 MetArgSerPro 195  
Db 3271 ATGAGGTCCCGG 3282

```
RESULT 13
ARI18728
LOCUS ARI18728 8987 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 137 from patent US 6150087.
ACCESSION ARI18728
VERSION ARI18728.1 GI:14100638
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 8987)
AUTHORS Chien,D.Y.
TITLE NANBV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 137 21-NOV-2000;
FEATURES
    source
        1..8987
        /organism="unknown"
BASE COUNT 1807 a 2735 c 2547 g 1898 t
ORIGIN

Alignment Scores:
Pred. No.: 1..3e-63 Length: 8987
Score: 894.50 Matches: 175
Percent Similarity: 89.71% Conservative: 8
Best Local Similarity: 85.78% Mismatches: 10
Query Match: 87.61% Indels: 11
DB: Gaps: 2

US-09-965-594-12 (1-195) x ARI18728 (1-8987)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 3013 CGCAGGCGCGGAGATGCTCGCGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 3072
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
Db 3073 TTGCTGGCGCCCATCATCGGCGTACGCCACGACAGAGGGGCTCTTAGGTGCATATC 3132
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 51
Db 3133 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTCAGTCCAGATTGTGTCAACT 3192
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGly 71
Db 3193 GCTGCCCAAACTTCCTCGCAACGTGCATCAATGGGCTGTGCTGGACTGTCTACCAAGG 3252
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 3253 GCGGACAGGAGGACCATCGCGTCACCCAAAGGTCCTGTCATCCAGATGTATACCAATGTA 3312
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 3313 GACCAACACCTTGTGGCTGGCGGCTCGCAAGGTAGCGGCTCATGTACACCTGCAC 3372
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 131
Db 3373 TCGCGCTCTCGGACCTTACCTGGTCAGGAGGACGCCGATGCTATCCCGTCGCGCGG 3432
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSortyrLeuLysGlySer 151
Db 3433 CGGGGTGATAGCAGGCGGACGCTGCTGCGCCCGCGCCATTCCTACTTTGAAGGCTCC 3492
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 3493 TCGGGGGTCCGCTGTTGTGCGCGCGGCGGACGCCGTCGGCATATTAGGGCGCGGTG 3552
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 3553 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3612
QY 192 MetArgSerPro 195
Db 3613 ATGAGGTCCCGC 3624

US-09-965-594-12 (1-195) x ARI18722 (1-9185)
QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17
Db 3332 CGCAGGCGCGGAGATGCTCGCGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 3391
QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 31
Db 3392 TTGCTGGCGCCCATCATCGGCGTACGCCACGACAGAGGGGCTCTTAGGTGCATATC 3451
QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 51
Db 3452 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTCAGATTGTGTCAACT 3511
QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValIleHisGly 71
Db 3512 GCTGCCCAAACTTCCTCGCAACGTGCATCAATGGGCTGTGCTGGACTGTCTACCAAGG 3571
QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91
Db 3572 GCGGACAGGAGGACCATCGCGTCACCCAAAGGTCCTGTCATCCAGATGTATACCAATGTA 3631
QY 92 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 111
Db 3632 GACCAAGACCTTGTGGCTGGCGGCTCGCAAGGTAGCGGCTCATGTACACCTGCAC 3691
QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 131
Db 3692 TCGCGCTCTCGGACCTTACCTGGTCAGGAGGACGCCGATGCTATCCCGTCGCGCGG 3751
QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSortyrLeuLysGlySer 151
Db 3752 CGGGGTGATAGCAGGCGGACGCTGCTGCGCCCGCGCCATTCCTACTTTGAAGGCTCC 3811
QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
Db 3812 TCGGGGGTCCGCTGTTGTGCGCGCGGCGGACGCCGTCGGCATATTAGGGCGCGGTG 3871
QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
Db 3872 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3931
QY 192 MetArgSerPro 195
Db 192 MetArgSerPro 195
```

db 5254 ATGAGGTCCCCG 5243

Search completed: August 31, 2003, 00:45:40  
Job time : 2558.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 181.082 seconds  
(without alignments)  
2905.924 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKGSWIVGRVINGAYA.....VAKAVDFIPVESLETTMRSP 195

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=rlp  
-Q/cgn2\_1/USPTO.spool/US09965594/runat\_29082003\_151918\_28302/app\_query.fasta\_1.2872  
-DB=N\_Geneseq\_19Jun03 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humaa40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-NO\_MMAPP -LARGESEQ=0 -CGN\_1\_1412=runat\_29082003\_151918\_28302 -NCPU=6 -ICPU=3  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.:

- 1: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/gnecseq/gnecseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1021	100.0	588	21	AAA73329 Hepatitis C virus
2	998	97.7	594	21	AAA73330 Hepatitis C virus
3	982	96.2	588	21	AAA73328 Hepatitis C virus
4	981	96.1	594	21	AAA73331 Hepatitis C virus
5	959	93.9	594	21	AAA73335 Hepatitis C virus
6	951	93.1	594	21	AAA73332 Hepatitis C virus
7	939	92.0	594	21	AAA73333 Hepatitis C virus
8	929	91.0	594	21	AAA73334 Hepatitis C virus
9	928.5	90.9	12734	24	ABA95615 Chimeric BVDV/HCV
10	918.5	90.0	612	25	ABX15706 Anti-viral synthet
11	894.5	87.6	5300	10	AAH92097 Combined open read
12	894.5	87.6	5360	10	AAH90327 Hepatitis C virus
13	894.5	87.6	6905	10	AAH92103 Combined open read
14	894.5	87.6	7310	10	AAH92106 Combined open read
15	894.5	87.6	7310	10	AAH90336 Composite hepatiti
16	894.5	87.6	7310	16	AAH98221 Hepatitis C virus
17	894.5	87.6	8316	21	AAH75296 cDNA sequence comp
18	894.5	87.6	9133	20	AAZ07656 Nucleotide sequenc
19	894.5	87.6	9185	11	AAQ05956 Sense strand of th
20	894.5	87.6	9185	12	AAQ10566 Hepatitis C virus
21	894.5	87.6	9185	21	AAH75297 Sense strand of HC
22	894.5	87.6	9401	13	AAQ21744 Compiled HCV cDNA.
23	894.5	87.6	9401	17	AAH12710 Hepatitis C virus
24	894.5	87.6	9401	18	AAH99981 HCV polyprotein co
25	894.5	87.6	9401	19	AAV09989 HCV polyprotein co
26	894.5	87.6	9401	24	AAD35043 Hepatitis C virus
27	894	87.6	549	21	AAA70344 Hepatitis C virus
28	894	87.6	2058	24	AAD29795 HCV-1 NS3/4a mutan
29	894	87.6	2058	24	ABK15344 Hepatitis C virus
30	894	87.6	2058	25	ABK14410 DNA encoding HCV-1
31	893.5	87.5	9502	15	AAO74770 Hepatitis C virus
32	892	87.4	1933	20	AAH23258 HCV NS3 DNA. Hepa
33	892	87.4	8145	20	AAH23259 Plasmid pET-BS(+)/
34	891.5	87.3	1998	20	AAH80355 HCV NS4A-NS3 compl
35	891.5	87.3	9185	20	AAH26737 Nucleotide sequenc
36	891.5	87.3	9185	20	AAH00459 Hepatitis C virus
37	890.5	87.2	8316	11	AAQ05955 Hepatitis C virus
38	890	87.2	2061	24	AAD34500 Hepatitis C virus
39	890	87.2	2061	24	AAD31767 Hepatitis C virus
40	889.5	87.1	9646	19	AAV59361 Hepatitis C virus
41	889.5	87.1	9646	24	ABK87285 cDNA encoding hepa
42	889.5	87.1	12980	19	AAV59364 Hepatitis C virus
43	889.5	87.1	12980	24	ABK87286 Hepatitis C virus
44	889.5	87.1	16622	21	AAH36212 Nucleotide sequenc
45	888.5	87.0	1998	20	AAH80359 HCV NS4A-NS3 compl

ALIGNMENTS

RESULT 1	
AAA73329	
ID	AAA73329 standard; DNA; 588 BP.
XX	
AC	AAA73329;
XX	
DT	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.
XX	
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW	liver failure; liver cancer; mutant; mutain; ds.
OS	Hepatitis C virus.
OS	Synthetic.
XX	
PH	Key Location/Qualifiers





```
XX SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;
Alignment Scores:
Pred. No.: 5, 64e-82 Length: 594
Score: 998.00 Matches: 193
Percent Similarity: 98.48% Conservativity: 1
Best Local Similarity: 97.97% Mismatches: 1
Query Match: 97.75% Indels: 2
DB: 21 Gaps: 1

US-09-965-594-12 (1-195) x AAA73330 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18
DQ 1 ATGAAAAAAGAGTCCGTTTATCGTCGGCCGTATCAACCTGTCGGTGACACCGCT 60
QY 19 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 38
DQ 61 TAGGCTCAGCAGCTCGAGGTGAGGAGGTTCCTCAAGAAACCTCCAGACCGGTGCTGAC 120
QY 39 LysAsnGlnValGluGlyGlnValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
DQ 121 AAAACACAGTTGAAGTGAAGTTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTGCGT 180
QY 59 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 78
DQ 181 ACCTGTCATCAACGGTGTTCGTGGACCGTTTACCACGGTGTGCTGCTACCGTACCACGCT 240
QY 79 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysLeuValGlyTrp 98
DQ 241 TCCCGAAGGTCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGTTGTTGG 300
QY 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
DQ 301 CGGCTCCGACGGTTCCTGCTGACCGCTGCACCTGCGGTTCCTCCACCTGTAC 360
QY 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
DQ 361 CTGGTTACCGTCACGCTGACGTTATCCCGTTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 158
DQ 421 CTGCTGTCCCGCGTCGTCATCTCCACCTGAAGGTTCTCCGTTGCTGCTGCTGCTGCT 480
QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
DQ 481 CCGGCTGTCACGCTGTCGTCATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540
QY 179 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
DQ 541 GCTGTTGACTTATCCCGGTTGAATCCCTGGAACACCATCGCTTCCCGC 591

RESULT 3
AAA73328
ID AAA73328 standard; DNR: 588 BP.
AC AC
XX XX
XX 19-DEC-2000 (first entry)
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.
KW Hepatitis C virus NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; ds.
XX Hepatitis C virus.
OS Synthetic.
XX Key
FH Location/Qualifiers
FT 1..588
FT CDS /tag= a
FT /product= "NS3-NS4A fusion protein"
```

Db		361	ACCCGTCAGCTGACGTATCCCGTTCGGTCGTCGGTGACTCCCGTGGTTCCCTGC	TG 421
QY		141	SerProArgPtoIleSerTyrLeuLySGLySerSerGlyGlyProLeuLeuCysProAla	160
Db		421	TCCC CGCGTCGATCTCCTACCTCAAAAGGTTCCTCCGTCGTGTCGTCGCGGCT	480
QY		161	GlyHisAlaValGlyLlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaVal	180
Db		481	GTCTACCGTGTGGGTATCTCCGTCGTCGTTTGCACCGTCGTGTGCTAAAGCTGT	540
QY		181	AspPheIleProValGlusErLeuGluthrThrMetArgSerPro	195
Db		541	GACTTCATCCCGGTGAATCCCTGMAAACCCACCATGCGTTC	585
<b>RESULT 4</b>				
AAA73331		ID	AAA73331 standard; DNA; 594 BP.	
XX		AC	AAA73331;	
XX		DT	19-DEC-2000 (first entry)	
XX		DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.	
KW		KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;	
KW		KW	Liver failure; liver cancer; mutant; mutein; ds.	
OS		OS	Hepatitis C virus.	
OS		OS	Synthetic.	
XX		Key	Location/Qualifiers	
FH		CDS	1..594	
FT		FT	/tag= a	
FT		FT	/product= "NS4A-NS3 fusion protein #4"	
XX		WO	2000040707-A1.	
PN		13	JUL-2000.	
XX		06	JAN-2000; 2000WO-US00345.	
PF		08	JAN-1999; 99US-0115271.	
XX		(BRIM )	BRISTOL-MYERS SQUIBB CO.	
PA		Wittekind M,	Weinheimer S, Zhang Y, Goldfarb V;	
PI		WPI; 2000-465976/40.	P-PSDB; AAB15222.	
DR		Modified hepatitis C virus (HCV) NS3 protease comprising at least 1	substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic	
XX		amino acid, useful for screening inhibitors that may treat hepatitis C	.	
PT		Claim 26; Fig 14; 66pp; English.		
XX		The present sequence is the coding sequence for a mutated version of a	fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A	
CC		protease enzymes. These proteins are both essential for the replication	of the virus, acting to cleave its replicative proteins from the	
CC		polyprotein produced from the HCV genome. Inhibitors of the two proteins	should be effective as antiviral treatments of HCV infection. This is	
CC		useful as HCV can lead to chronic liver disease such as cirrhosis, liver	failure and liver cancer. The present invention concerns a number of NS3	
CC		mutants and NS3-NS4A fusion proteins which can be used to identify	inhibitors of this type, as well as enabling structural studies of the	
CC		protease and protease-inhibitor complexes. The protein produced from this	sequence contains the alpha-heli0-1 variant.	
XX		Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other:		
SQ		Alignment Scores:		

Pred. No.:	1.99e-80	Length:	594
Score:	981.00	Matches:	190
Percent Similarity:	96.95%	Conservative:	1
Best Local Similarity:	96.45%	Mismatches:	4
Query Match:	96.08%	Indels:	2
DB:	21	Gaps:	1
US-09-965-594-12 (1-195) * AAA73331 (1-594)			
QY	1	MetLysLysGlySerValValIleValGlyArgIleValLeuLeuAsnGly-----Ala	18
DB	1	ATGAAAAAAAAGGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCCGGTGACACCGCT	60
QY	19	TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp	38
DB	61	TAGGCTCAGCAGACTCGAGGTGAGAGGGTTGCCAAGAAACCTCCCAAGACCGGTCTGTGAC	120
QY	39	LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla	58
DB	121	AAAAACCAAGTTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT	180
QY	59	ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla	78
DB	181	ACCTGCATCAACGGTGTTCGTGGACCGCTTTACCACGGTGTGTGTACCCGATCGCT	240
QY	79	SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr	98
DB	241	TCCCGAAGGTCGCGTTACCCAGATGTACACCAACGTTGCACAAGACCTGGTTGGTTGG	300
QY	99	ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr	118
DB	301	CAGGCTCCGACGGGTCCCGTTCCCTGACCCGCTGCACCTCGGTTCTCCGACCTGTAC	360
QY	119	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer	138
DB	361	CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTGCTGCTGTGTGACTCCCGGTTC	420
QY	139	LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys	158
DB	421	CTGCTCTCCCGCGCTCCGATCTCCCTACCTGAAAGGTTCTCCGCTGCTGCTGCTGC	480
QY	159	ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys	178
DB	481	CGGGTGGTTCAGCTGTGGTATCTTCGCTGCTGTTCGACCCGCTGGTGTGTCTAA	540
QY	179	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	195
DB	541	GCTGTTGACTTCATCCGGTTGAATCTCCCTGGAACACCATGCTTCCCCG	591
RESULT 5			
AAA73335			
ID	AAA73335 standard; DNA; 594 BP.		
XX	AAA73335;		
AC			
XX			
DT	19-DEC-2000 (first entry)		
XX			
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.		
XX			
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;		
KW	liver failure; liver cancer; mutant; mutein; ds.		
XX			
OS	Hepatitis C virus.		
OS	Synthetic.		
XX			
XX	Key Location/Qualifiers		
FT	1..594		
FT	/*tag= a		
FT	/product= "NS4A-NS3 fusion protein #8"		
XX			
PN	W0200040707-A1.		
XX			
PD	13-JUL-2000.		

```
XX 06-JAN-2000; 2000NO-US00345.
XX 08-JAN-1999; 9905-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
XX P-PSDB; AAB15226.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
XX Disclosure; Fig 18; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0 wild-type sequence.
XX
XX Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2e-78 Length: 594
XX Score: 959.00 Matches: 188
XX Percent Similarity: 95.94% Conservativeness: 1
XX Best Local Similarity: 95.43% Mismatches: 6
XX Query Match: 93.93% Indels: 2
XX DB: 21 Gaps: 1
XX
XX
XX US-09-965-594-12 (1-195) x AAA73335 (1-594)
XX
XX 1 MetLysLysGlySerValValIleValGlyArgIleValLeuAAsngly-----Ala 18
XX
XX 1 ATGAAAGGATCGGTTGTTATCGTCGGCGGTATCAACCTGTCCGGTGACACCGCT 60
XX
XX 19 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 38
XX
XX 61 TACGCTCAGCAGACTCGAGGTCCTCGGTGGATCATCATCCTCCCTGACCGGTCGTGAC 120
XX
XX 39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
XX
XX 121 AAAAACAGGTTGAAGGTGAAGTTTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTCGCT 180
XX
XX 59 ThrCysIleAsnGlyValCysTrpThrValTrpHisGlyAlaGlyThrArgThrIleAla 78
XX
XX 181 ACCTGCATCAACGGTGTGTGACCGGTTTACCACGGTGTGTGTACCGGTACCATCGCT 240
XX
XX 79 SerProLysGlyProValIleGlnMetThrThrAsnValAspLysAspLeuValGlyTrp 98
XX
XX 241 TCCCGGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTTGG 300
XX
XX 99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
XX
XX 301 CCGGCTCCGACGGTTCCCGTTCCTGACCCCGTGCACCTCGCGTTCTCCGACCTGTAC 360
XX
XX 119 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 138
XX
XX 361 CTGGTTACCCGCTCAGGCTGACGTTATCCCGGTTCTGCTCGTGTGACCTCCCGTGGTTCC 420
XX
XX 139 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 158
```

```
Db 421 CTGCTGTCCCGCGCCGATCTCCTACCTGAAGGTTCTCCGGTGGTCCGCTGCTGTGC 480
QY 159 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 178
Db 481 CCGGCTGTGTCACCGCTGGTGTATCTTCGCTGCTGCTGTTGCACCGGTGGTGGCTAAA 540
QY 179 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 195
Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTCGGAACACCATGCGTCCCG 591
RESULT 6
AAA73332
ID AAA73332 standard; DNA; 594 BP.
XX
XX AAA73332:
XX
XX 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutein; ds.
XX
XX Hepatitis C virus.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..594
XX FT /tag= a
XX FT /product= "NS4A-NS3 fusion protein #5"
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000NO-US00345.
XX
XX 08-JAN-1999; 9905-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
XX P-PSDB; AAB15223.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
XX Claim 26; Fig 15; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
XX Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.07e-77 Length: 594
XX Score: 951.00 Matches: 187
XX Percent Similarity: 95.43% Conservativeness: 1
```

	Best Local Similarity:	94.92%	Mismatches:	7
Query Match:	93.14%	Indels:	2	
DB:	21	Gaps:	1	

US-09-965-594-12 (1-195) x AAA73332 (1-594)

QY	1	MetLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18 
Dd	1	ATGAATAAAAGAGTCCGTGGTTTCCTCGGCCGTATCAACCTGTCGGTGACACCGCT 60 
QY	19	TyrAlaGlnThrArgGlyCyluGlyCysGlnGluThrSerGlnThrGlyArgAsp 38 
Dd	61	TAGGCTCACGACTCGAGTGCAGAGGTTGCCAAGAACCCTCCAGACGGTCTGTGAC 120 
QY	39	LysAsnGlnValGlucyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58 
Dd	121	AAAAAACAGGTGAAGTGAAGTTTCAGATCGTTCCACCGCTACCCAGACCTCTCGTGGT 180 
QY	59	ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78 
Dd	181	ACCTCATCAACGGTCTCTGTGGACCGTTTTACACGGGTGCTGGTAGCCATTCTGCTG 240 
QY	79	SerProLysGlyProValIleGlnMetTyThrAsnValAspLysAspLeuValGlyTrp 98 
Dd	241	TCCCCGAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAGACCTGGTGGTGG 300 
QY	99	ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118 
Dd	301	CAGGCTCCGACGGTTCCTGCTCCGTACCCCGTGCACCTCGCTCTCCGACCTGTAC 360 
QY	119	LeuValThrArgHisAlaAspVallieProValArgArgGlyAspSerArgGlySer 138 
Dd	361	CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCTGCTCGTGGTGAACCTCGGTTC 420 
QY	139	LeuLeuSerProArgProfileSertyrLeuLysGlySerSerGlyGlyProLeuLeuCys 158 
Dd	421	CTGCTGTCCC CGCGCCGATCCTACCTGAAAGGTTCTCCGGTGGTCCGCTGCTGTC 480 
QY	159	ProAlaGlyHisAlaValcylePheArgAlaAlaValCysThrArgGlyValAlaLys 178 
Dd	481	CCGGCTGTGCAGCTGTGGTATCTCCGTCTGCTGTCTCCACCCGTGGTGTGCTAAA 540 
QY	179	AlaValAspPheileProValGluSerLeuGluThrThrMetArgSerPro 195 
Dd	541	GCTGTGTAGTTTCATCCCGGTTGAATCCCTGGAACCCACCATCGCTCCCG 591 

RESULT 7  
AAA73333  
ID AAA73333 standard; DNA; 594 BP.

XX	AA	AAA73333;
AC	XX	AAA73333;
XT	19-DEC-2000	(first entry)
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.	
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease; liver failure; liver cancer; mutant; mutein; ds.	
OS	Hepatitis C virus.	
FT	Synthetic.	
PH	key Location/Qualifiers	
CD	1..594 /tag= a	
FT	/product= "NS4A-NS3 fuslon protein #6"	
PB	WO200040707-A1.	
PD	13-JUL-2000.	
PF	06-JAN-2000; 2000MO-US00345.	
XX		

```

PR      08-JAN-1999;   99US-0115271.
XX      (BRIM ) BRISTOL-MYERS SQUIBB CO.
PA      Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX      WPI: 2000-465976/40.
DR      P-PADB; AAB15224.
XX
PT      Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT      substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT      amino acid, useful for screening inhibitors that may treat hepatitis C
PT      .
XX
PS      Claim 26; Fig 16; 6pp; English.
XX
CC      The present sequence is the coding sequence for a mutated version of a
CC      fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC      protease enzymes. These proteins are both essential for the replication
CC      of the virus, acting to cleave its replicative proteins from the
CC      polypotein produced from the HCV genome. Inhibitors of the two proteins
CC      should be effective as antiviral treatments of HCV infection. This is
CC      useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC      failure and liver cancer. The present invention concerns a number of NS3
CC      mutants and NS3-NS4A fusion proteins which can be used to identify
CC      inhibitors of this type, as well as enabling structural studies of the
CC      protease and protease-inhibitor complexes. The protein produced from this
CC      sequence contains the alpha-helix0-7 variant.
XX
SQ      Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:
Pred. No.:    1.32e-76          Length:        594
Score:         939.00           Matches:       184
Percent Similarity:  94.92%     Conservative:   3
Best Local Similarity: 93.40%    Mismatches:    8
Query Match:    91.97%          Indels:       2
DB:             21              Gaps:         1

US-09-965-594-12 (1-195) x AAA73333 (1-594)

QY      1 MetLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala 18
Dd      1 ATGAATAAAAGAGTCCGTGGTTTCCTCGGCCGTATCAACCTGTCGGTGACACCGCT 60
QY      19 TyrAlaGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 38
Dd      61 TAGGCTCACGACTCGAGTGCAGAGGTTGCCAAGAACCCTCCACACGGTCTGTGAC 120
QY      39 LysAsnGlnValGlucyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
Dd      121 AAAAAACAGGTGAAGTGAAGTTTCAGATCGTTCCACCGCTACCCAGACCTCTCGTGGT 180
QY      59 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78
Dd      181 ACCTCATCAACGGTCTCTGTGGACCGTTTTACACGGGTGCTGGTAGCCATTCTGCTG 240
QY      79 SerProLysGlyProValIleGlnMetTyThrAsnValAspLysAspLeuValGlyTrp 98
Dd      241 TCCCCGAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAGACCTGGTGGTGG 300
QY      99 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 118
Dd      301 CAGGCTCCGACGGTTCCTGCTCCGTACCCCGTGCACCTCGCTCTCCGACCTGTAC 360
QY      119 LeuValThrArgHisAlaAspVallieProValArgArgGlyAspSerArgGlySer 138
Dd      361 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCTGCTCGTGGTGAACCTCGGTTC 420
QY      139 LeuLeuSerProArgProfileSertyrLeuLysGlySerSerGlyGlyProLeuLeuCys 158
Dd      421 CTGCTGTCCC CGCGCCGATCCTACCTGAAAGGTTCTCCGGTGGTCCGCTGCTGTC 480
QY      159 ProAlaGlyHisAlaValcylePheArgAlaAlaValCysThrArgGlyValAlaLys 178
Dd      481 CCGGCTGTGCAGCTGTGGTATCTCCGTCTGCTGTCTCCACCCGTGGTGTGCTAAA 540

QY      179 AlaValAspPheileProValGluSerLeuGluThrThrMetArgSerPro 195
Dd
Dd      541 GCTGTGTAGTTTCATCCCGGTTGAATCCCTGGAACCCACCATCGCTCCCG 591

RESULT 7
AAA73333
ID AAA73333 standard; DNA; 594 BP.
XX AA
XX AAA73333;
AC XX
AC AAA73333;
XT 19-DEC-2000 (first entry)
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
XX OS
XX OS Hepatitis C virus.
XX FT
XX PH key Location/Qualifiers
XX CD 1..594 /tag= a
XX FT /product= "NS4A-NS3 fuslon protein #6"
XX PB WO200040707-A1.
XX PD 13-JUL-2000.
XX PF 06-JAN-2000; 2000MO-US00345.
XX
XX
XX

```

QY	159	PROAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys	179
DB	481	CGCGTGGTCACGCTGTTGGTATCTCCGTCGTCGTTTCCACCGTGGTGTGCTAA	540
QY	179	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	195
DB	541	GCTGTGTACTTCATCCCGGTGAATCCTGGAAACCACCATGCTTCCCG	591
RESULT 8			
AAA73334			
ID	AAA73334 standard; DNA; 594 BP.		
XX	AC	AAA73334;	
DT	19-DEC-2000 (first entry)		
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.		
XX	Hepatitis; NS3 protease; viral replication; chronic liver disease;		
KW	liver failure; liver cancer; mutant; mutein; ds.		
XX	Hepatitis C virus.		
OS	Synthetic.		
Key	Location/Qualifiers		
FF	1..594		
FT	/*tag= a		
FT	/product= "NS4A-NS3 fusion protein #7"		
XX	WO200040707-AL.		
PN	13-JUL-2000.		
PD	06-JAN-2000; 2000WO-US00345.		
PF	08-JAN-1999; 99US-0115271.		
XX	(BRIM ) BRISTOL-MYERS SQUIBB CO.		
PA	Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;		
PI	WPI; 2000-465976/40.		
DR	P-PSDB; AAB15225.		
XX	Modified hepatitis C virus (HCV) NS3 protease comprising at least 1		
PT	substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic		
PT	amino acid, useful for screening inhibitors that may treat hepatitis C		
PT	-		
XX	Claim 26; Fig 17; 66pp; English.		
XX	The present sequence is the coding sequence for a mutated version of a		
CC	fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A		
CC	protease enzymes. These proteins are both essential for the replication		
CC	of the virus, acting to cleave its replicative proteins from the		
CC	polyprotein produced from the HCV genome. Inhibitors of the two proteins		
CC	should be effective as antiviral treatments of HCV infection. This is		
CC	useful as HCV can lead to chronic liver disease such as cirrhosis, liver		
CC	failure and liver cancer. The present invention concerns a number of NS3		
CC	mutants and NS3-NS4A fusion proteins which can be used to identify		
CC	inhibitors of this type, as well as enabling structural studies of the		
CC	protease and protease-inhibitor complexes. The protein produced from this		
CC	sequence contains the alpha-helix0-7 variant.		
XX			
SQ	Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;		
Alignment Scores:			
Pred. No.:		1..08e-75	Length: 594
Score:		929.00	Matches: 183
Percent Similarity:		94.42%	Conservative: 3
Best Local Similarity:		92.89%	Mismatches: 9
Query Match:		90.99%	Indels: 2
DB:		21	Gaps: 1

US-09-965-594-12 (1-195) x AAA73334 (1-594)

QY	1	MetLysLysLysGlySerValValIleValGlyArgIleValLeuAsnGly-----Ala18
Db	1	ATGAAAAAAGAGATCGTTGTTATCGTCGGCGGTATCAACCTGTCCCGTGACACCGCT60
QY	19	TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp38
Db	61	TACGCTCAGCAGACTCGAGGTGACAGGGTACCCAGAGACCTCCACACACCGGTCTGTGAC120
QY	39	LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla58
Db	121	AAAAACAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT180
QY	59	ThrCysIleAsnGlyValCysTrpThrValTrpHisGlyAlaGlyThrArgThrIleAla78
Db	181	ACCTCCATCAACGGGTGTTCTGTGGACCGTTTACCACACGGTCTGTGTACCCGTACCATCGCT240
QY	79	SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp98
Db	241	TCCCCGAAGGTCCGGTTACCCAGATGTTACCCAAACGTTGACAAAGACCTGGTGGTTGG300
QY	99	ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr118
Db	301	CAGGCTCCGAGGTTCCCGTCTCCAGCCCGTGACCTGCGGTCTCCGACCTGTAC360
QY	119	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer138
Db	361	CTGTTTACCCGTCACGCTGAGCTTATCCGGTTCGTGCTGTGTGAGTCTCCCGTGGTTCC420
QY	139	LeuLeuSerProArgProIleSerTyrIleuLysGlySerSerGlyGlyProLeuLeuCys158
Db	421	CTGCTGTCCCGCGTCCGATCTCCTACCTGAAGAGTTCTCTCCGCTGTGCTGGCTGTGC480
QY	159	ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys178
Db	481	CCGGCTGGTCACGCTGGTGGTATCTTCGCTGCTGTGTTCACCCCGTGGTGTGCTAA540
QY	179	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro195
Db	541	GCTGTTGACTTCATCCGCTTGAATCCCTGGAAACCAACCATGCGTTCGCCG591
RESULT 9		
ABA95615		
ID	ABA95615	standard; DNA; 12734 BP.
XX	ABA95615;	
DT	21-MAR-2002	(first entry)
DE	Chimeric BVDV/HCV NS3-wt sequence.	
XX	Pestivirus; Npro; protease; NS3; screening; ds.	
XX	Chimeric - Bovine viral diarrhoea virus.	
XX	Chimeric - Hepatitis C virus.	
XX	US6326137-B1.	
XX	04-DEC-2001.	
XX	25-JUN-1999;	99US-0344456.
XX	25-JUN-1999;	99US-0344456.
XX	(SCHE ) SCHERING CORP.	
XX	Hong Z, Lai VCH, Lau JYN;	
XX	WPI; 2002-121103/16.	
XX	Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)	

PT pestivirus genome where the Npro protease gene is replaced with NS3  
PT protease gene, useful for in vivo screening of compounds which inhibit  
XX HCV infection -  
XX  
XX Example 2; Columns 17-28; 20pp; English.  
XX  
CC The present invention relates to a nucleic acid construct encoding a  
CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
CC each junction site recognised by the Npro protease is replaced with a  
CC junction site recognised by the HCV NS3 protease. The construct is useful  
CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
CC protease, where screening may be in cell culture or in an animal model.  
CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
CC virus)/HCV NS3-wt, which was used to illustrate the present invention.  
XX  
SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.05e-74 Length: 12734  
Score: 928.50 Matches: 180  
Percent Similarity: 94.87% Conservativity: 5  
Best Local Similarity: 92.31% Mismatches: 5  
Query Match: 90.94% Indels: 5  
DB: 24 Gaps: 1  
  
US-09-965-594-12 (1-195) x ABA95615 (1-12734)  
QY 5 GlySerValIleValGlyArgIleValLeuAsnGly-----AlaTyr 19  
DB 413 GGTAGTGTCTTATTTGGTAGAATTTTATCTGGTAGTGGTAGTACACGGCGTAC 472  
  
QY 20 AlaGlnThrArgGlyGluGluGlyCysGlnThrSerGlnThrGlyArgAspLys 39  
DB 473 GCCAGCAGAGAGAGCCCTCTAGGGTGTAGATCACCAGTCTGACTGGCCGGACAAA 532  
  
QY 40 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 59  
DB 533 AACCAAGTGGAGGTGAGGTGCCAGATGCTGCTCACTACCCAAACCTTCTCGCAACG 592  
  
QY 60 CysIleAsnGlyValCysTpeThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 79  
DB 593 TGCATCAATGGGTATGCTGGAGTGTCTACACGGGCGGGAACGAGGACCATCGCATCA 652  
  
QY 80 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro 99  
DB 653 CCCAAGGTCTCTCATCCAGATGTATACCAATGTGGACCAACCTTGTGGCTGGCCC 712  
  
QY 100 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 119  
DB 713 GCTCCTCAAGGTTCCCGCTCATTGACACCTCGACCTGCGGCTCCTCGGACCTTTACCTG 772  
  
QY 120 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 139  
DB 773 GTTACAGGACGCGGACGTCATTCCTCGCGCGGAGGTGATACAGGGGTAGCGCTG 832  
  
QY 140 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 159  
DB 833 CTTTCGCGCGCGCCATTCTCTACCTAAAGGCTCTCTCGGGGGTCCGCTGTGTGCCCC 892  
  
QY 160 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 179  
DB 893 GCGGACACGCGGTGGGCTATTTCAGGCGCGGTGTGCACCGGTGGAGTGGCAAGGCG 952  
  
QY 180 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194  
DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGACACACCATGATGCC 997  
  
RESULT 10  
ABX15706  
ID ABX15706 standard; DNA: 612 BP.  
XX

AC ABX15706;  
XX  
XX 28-MAR-2003 (first entry)  
XX  
XX Anti-viral synthetic prototoxophore associated DNA sequence.  
XX  
XX Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;  
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;  
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.  
XX  
XX Unidentified.  
XX  
XX WO200287500-A2.  
XX  
XX 07-NOV-2002.  
XX  
XX 26-APR-2002; 2002WO-US13223.  
XX  
XX 27-APR-2001; 2001US-286893P.  
XX  
XX (NEWB-) NEWBIOTICS INC.  
XX  
XX Cathers BE, Neuteboom STC, Shepard HM;  
XX  
XX WPI; 2003-167102/16.  
XX  
XX Novel synthetic viral prototoxophore for treating viral infections, has  
PT toxin moiety incorporated into substrate domain specific for viral  
PT enzyme, bound and modified by viral enzyme to get converted into  
PT toxophore -  
XX  
XX Example 1; Page 62; 66pp; English.  
XX  
XX This invention relates to a novel synthetic viral prototoxophore  
CC comprising a toxin moiety operatively incorporated into a substrate  
CC domain specific for a viral enzyme. This prototoxophore may be bound  
CC and modified by the viral enzyme thus converting it to a toxophore.  
CC Also disclosed in the invention is a method for enhancing the anti-viral  
CC effect of an antiviral agent, this method comprises contacting a cell,  
CC infected with a virus or is susceptible to infection, with a  
CC prototoxophore. The invention further comprises an assay to identify  
CC anti-viral agents, comprising contacting an infected cell with a  
CC candidate agent and comparing the ability of the agent to inhibit the  
CC growth or infectivity of the virus in the cell. The prototoxophores  
CC of the invention may have virucide or antitumour activity. The  
CC prototoxophores of the invention may be useful for reducing or  
CC inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,  
CC nerve cell, connective tissue cell, muscle cell or hepatocyte) which is  
CC infected with a virus or is susceptible to infection with a virus, with  
CC an effective amount of the prototoxophore. The cells are cell lines  
CC adapted to long term continuous culture or isolated from a subject.  
CC The prototoxophore is also useful for ameliorating the severity of a  
CC viral infection in a subject, where the virus is selected from human  
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and  
CC hepatitis virus, by administering an effective amount of the  
CC prototoxophore to the subject. The prototoxophores of the invention are  
CC also useful for treating tumours. The present sequence represents an  
CC antiviral prototoxophore associated DNA sequence, this sequence is  
CC described as a recombinant NS3/NS4 fusion protein in example 1 of  
CC the invention although it is clearly not a protein sequence.  
XX  
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1.01e-74 Length: 612  
Score: 918.50 Matches: 179  
Percent Similarity: 94.36% Conservativity: 5  
Best Local Similarity: 91.79% Mismatches: 6  
Query Match: 89.96% Indels: 5  
DB: 25 Gaps: 1  
  
US-09-965-594-12 (1-195) x ABX15706 (1-612)

QY 5 GlySerValValIleValGlyArgIleValLeuAsnGly-----AlaTyr 19  
 |||||  
 Db 19 GSTAGTGGTCAATTGGTGGTAGGATCAATTTGTCGGTAGGGAGTATCAACGGCTAC 78  
 QY 20 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 39  
 |||||  
 Db 79 GCCACGACAGCAAGGGGCTCTAGGTGTCATATCAACAGCCTAACTGGCGGGACAAA 138  
 QY 40 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 59  
 |||||  
 Db 139 AACCAAGTGGAGGTGAGGTGCAGATTGTGTCACTGCTGCCAAACCTTCTGTGCAACG 198  
 QY 60 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 79  
 |||||  
 Db 199 TGCAATCAATGGGGTGGTGGGACTGTCTACACGGGGCGGGAACAGACCATCGCTCA 258  
 QY 80 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 99  
 |||||  
 Db 259 CCCAAGGTCTGTATCCAGATGTATACCAATGTAGACCAAGACTTGTGGCTGGCCC 318  
 QY 100 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerAspLeuTyrLeu 119  
 |||||  
 Db 319 GCTTCCCAAGGTACCGCTCATTCACACCTGCACCTTGGCTGCTCGACCTTTACCTG 378  
 QY 120 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 139  
 |||||  
 Db 379 GTACAGGACGACGGATGTCTATCCGTGCGCGGGGGGTAGACAGGGGCGACCTG 438  
 QY 140 LeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCysPro 159  
 |||||  
 Db 439 CTGTGCGCCCGCCCATTTCTCTACTTGAAGGCTCTCGGGGGTCCGCTGTGTGGCCC 498  
 QY 160 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 179  
 |||||  
 Db 499 GCGGGGACGCGCGTGGCATATTTAGGCGCGGGTGTGCACCGGTGGAGTGGCTAAGCG 558  
 QY 180 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 194  
 |||||  
 Db 559 GTGGACTTATCCCTGTGGAGAACCTAGACAGAACCATGAGGTCC 603  
 RESULT 11  
 AAN92097  
 ID AAN92097 standard; DNA; 5300 BP.  
 XX  
 AC AAN92097;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX  
 DE Combined open reading frames of the hepatitis C virus (HCV) cDNA in  
 DE clones 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c,  
 DE 8f, 33f, 33g and 39c.  
 XX  
 KW Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 3..5300  
 FT /\*tag= a  
 XX  
 EP318216-A.  
 XX  
 PD 31-MAY-1989.  
 XX  
 PF 18-NOV-1988; 88EP-0310922.  
 XX  
 PR 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.

XX (CHIR ) CHIRON CORP.  
 XX Houghton M, Choo QL, Kuo G;  
 XX WPI: 1989-159274/22.  
 DR P-PSDB: AAP92041.  
 XX  
 PT Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 XX  
 PS Claim 3; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.  
 XX  
 CC It is a double-stranded nucleotide sequence of the open reading frame  
 CC (ORF) (tag a) extending through clones 14i, 11b, 7f, 7e, 8h 33c, 40b,  
 CC 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c of hepatitis C  
 CC virus (HCV) cDNA. In creating the composite sequence the following  
 CC heterogeneities were considered. Clone 33c contains a sequence  
 CC of 800 base pairs which overlaps the cDNAs in clones 40b and 37c. In  
 CC clone 33c, as well as in 5 other overlapping clones, nucleotide #789 is  
 CC a G. However, in clone 37b the corresponding nucleotide is an A. This  
 CC heterogeneity may have important ramifications for protein folding.  
 CC Nucleotide #2 in clone 8h is a T which may represent a cloning artifact  
 CC because the corresponding residue in clone 7e and in 3 other overlapping  
 CC clones is an A. Therefore the residue in this position is designated as  
 CC an A. The 3'-terminal nucleotide in clone 8f is represented as a T  
 CC than a G because the corresponding residue in clone 33f and in 2 other  
 CC overlapping clones is a T. The 3' terminal sequence of clone 33f is  
 CC represented as ATTC, as is found in the corresponding sequence in clone  
 CC 33g and in 2 other overlapping clones, rather than as TTGC, as is found  
 CC in clone 33f. Residue #4 in clone 33g is designated an A rather than a T  
 CC because the corresponding residue in clone 33f and 2 other overlapping  
 CC clones is an A. The 3'-terminus of clone 14i is depicted as TA rather  
 CC than AA because the corresponding dinucleotide in clone 11b and 3 other  
 CC clones is TA. Potential cloning artifacts have been omitted and instead  
 CC the corresponding sequences in non-5'-terminal regions of multiple  
 CC overlapping clones are shown. AAN92097 could be used as a source of  
 CC oligomeric DNA hybridisation probes to detect the presence of HCV  
 CC nucleic acids in samples. The polypeptide(s) it encodes could be used as  
 CC immuno- assay reagents and vaccines and to generate antibodies useful in  
 CC diagnosis and passive immunotherapy for HCV infection/non-A, non-B  
 CC hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX SQ Sequence 5300 BP; 1047 A; 1606 C; 1515 G; 1130 T; 2 other;

Alignment Scores:  
 Pred. No.: 2,15e-71 Length: 5300  
 Score: 894.50 Matches: 175  
 Percent Similarity: 89.71% Conservative: 8  
 Best Local Similarity: 85.78% Mismatches: 10  
 Query Match: 87.61% Indels: 11  
 DB: 10 Gaps: 2  
 US-09-965-594-12 (1-195) x AAN92097 (1-5300)  
 QY 3 LysLysGlySerValIleValGly-----ArgIleValLeuAsnGly----- 17  
 |||||  
 Db 867 CGCAGGGGCGGGAGATACCTCTCGGGCCAGCCGATGGAATGCTCCAAAGGGGTGGAGG 926  
 QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31  
 |||||  
 Db 927 TTGCTGGCGGCCATCAGCGCGGTACGCCACAGCAAGGGGCGCTCTTAGGTGCATAATC 986  
 QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 51  
 |||||  
 Db 987 ACCAGCTTAAGTCCCGGGACAAACCAAGTGGGGGTGAGGTCCAGATGTGTCAACT 1046  
 QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
 |||||  
 Db 1047 GCTGCCCAAAACCTTCTCTGGCAACCTGCATCATCGGGGTGTGTGACTGTCTACACGGG 1106

QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
 Db 1107 GCGGAACGAGGACCATCGCGTCACCAAGGGTCTGTATCCAGATGTATACCAATGTA 1166  
 QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
 Db 1167 GACCAAGACCTTGTGGCTGGCGCGCTCCCAAGGTAGCGGCTCATGTACACCCCTGCAT 1226  
 QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
 Db 1227 TGGCGCTCCTCGGACCTTACCTGGTACAGGACGACGCGATGTCATTCGCGCGCG 1286  
 QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
 Db 1287 CGGGGTGATAGCAGGGGACGCTGTCTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346  
 QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
 Db 1347 TCGGGGGTCCGTGTGTGCCCGCGGGGACGCGCGTGGGCATATTTAGGGCGCGTG 1406  
 QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
 Db 1407 TGCACCGTGGAGTGGCTAAGCGGTGGACTTTATCCTGTGGAGAACCCTAGAGACAACC 1466  
 QY 192 MetArgSerPro 195  
 Db 1467 ATGAGGTCCCG 1478

RESULT 12  
 AAN90327  
 ID AAN90327 standard; cDNA: 5360 BP.  
 AC AAN90327;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 11-NOV-1989 (first entry)  
 XX  
 XX Hepatitis C virus composite probe.  
 XX Hepatitis C virus; composite cDNA; probe; vaccine.  
 KW  
 XX  
 XX Pan troglodytes.  
 OS  
 PH Location/Qualifiers  
 FT CDS 3..5360  
 FT /\*tag= a  
 XX  
 XX GB2212511-A.  
 XX 26-JUL-1989.  
 PF 18-NOV-1988; 88GB-0027024.  
 XX  
 PF 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 26-OCT-1988; 88US-0263584.  
 XX  
 XX (CHIR ) CHIRON CORPORATION.  
 XX  
 XX Houghton M, Choo QL, Kuo G;  
 PI  
 XX  
 DR WPI: 1989-215054/30.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment  
 PT of infection.

XX Disclosure; Fig. 26; 174pp; English.

XX The sequence shows the composite cDNA sequence derived from the aligned  
 CC hepatitis C virus (HCV) cDNA's in clones 141, 11b, 7f, 7e, 8h, 33c, 40b,  
 CC 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA

CC encodes antigens which react with antibodies in patients with non-A  
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to  
 CC synthesize polypeptides, which can be used to diagnose HCV-induced NANBH,  
 CC to raise antibodies for immunoassay or treatment, or to produce  
 CC vaccines. See also AAP90158, AAN90303-26, and AAN90328-36.  
 CC (updated on 25-MAR-2003 to correct PR field.)  
 XX

SQ Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;

Alignment Scores:  
 Pred. No.: 2,18e-71 Length: 5360  
 Score: 84.50 Matches: 175  
 Percent Similarity: 89.71% Conservative: 8  
 Best Local Similarity: 85.78% Mismatches: 10  
 Query Match: 87.61% Indels: 11  
 DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN90327 (1-5360)

QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
 Db 867 CGCAGGGCGGGGAGATACCTGCTCGGGCCAGCGCATGGAATGCTCTCCAAGGGGTGAGG 926  
 QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 31  
 Db 927 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACAAAGGGGCTCTAGGGTGCATAATC 986  
 QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 51  
 Db 987 ACCAGCTAACTGGCGCGGACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCACT 1046  
 QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 71  
 Db 1047 GCTGCCAAACCTTCTGTGCAACGTGCATCAATGGGTGTGTGGACTGTCTACACGGG 1106  
 QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
 Db 1107 GCGGGAACGAGGACCATCGCGTACCCAAAGGTCTGTCTATCCAGATGTATACCAATGTA 1166  
 QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
 Db 1167 GACCAAGACCTTGTGGCTGGCGCGCTCCGCAAGGTAGCGGCTCATTCACACCTGCACT 1226  
 QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
 Db 1227 TCGGGTCTCTCGGACCTTTACCTGTGTACAGGACGCGCATGTATCCCTGTGAAGGCTCC 1286  
 QY 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
 Db 1287 CGGGGTGATAGCAGGGGACGCTGTCTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346  
 QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
 Db 1347 TCGGGGGTCCGTGTGTGCCCGCGGGGACGCGGTGGGCATATTTAGGGCGCGTG 1406  
 QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
 Db 1407 TGCACCGTGGAGTGGCTAAGCGGTGGACTTTATCCTGTGGAGAACCCTAGAGACAACC 1466  
 QY 192 MetArgSerPro 195  
 Db 1467 ATGAGGTCCCG 1478

RESULT 13  
 AAN92103  
 ID AAN92103 standard; DNA; 6905 BP.

XX AAN92103;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX

DE Combined open reading frames of the hepatitis C virus (HCV) cDNAs from



DE clones 12f through 15e.  
 XX Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 3..6905  
 FT /\*tag= a  
 FT  
 XX  
 XX EP318216-A.  
 XX  
 PD 31-MAY-1989.  
 XX  
 XX 18-NOV-1988; 88EP-0310922.  
 XX  
 PR 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Houghton M, Choo QL, Kuo G;  
 XX  
 XX WPI: 1989-159274/22.  
 DR P-PSDB; AAP2047.  
 XX  
 XX Purified hepatitis C virus  
 FT - and associated nucleic acids and polypeptide(s)  
 XX  
 XX Claim 3; Figure 32-1 - 32-7; 139pp; English.  
 XX  
 CC It is a double-stranded nucleotide sequence of the open reading frame  
 CC (ORF) (tag a) extending through clones 12f to 15e of hepatitis C virus  
 CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to  
 CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)  
 CC it encodes could be used as immunoassay reagents and vaccines and to  
 CC generate antibodies useful in diagnosis and passive immunotherapy for  
 CC HCV infection/non-A, non-B hepatitis.  
 CC  
 CC (updated on 25-MAR-2003 to correct PR field.)  
 CC (updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 6905 BP; 1421 A; 2082 C; 1946 G; 1456 T; 0 other;

Alignment Scores:  
 Pred. No.: 2.97e-71 Length: 6905  
 Score: 894.50 Matches: 175  
 Percent Similarity: 89.71% Conservative: 8  
 Best Local Similarity: 85.78% Mismatches: 10  
 Query Match: 87.61% Indels: 11  
 DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN92103 (1-6905)  
 QY 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
 Db 1140 CGCAGGGGGGGAGATATCTGCTGGGCCAGCGGATGGATGCTCCCAAGGGGTGGAGG 1199  
 QY 18 -----AlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31  
 Db 1200 TTGCTGGGCCCATCAGGGGTACGCCAGCACAAAGGGGCTCTAGGGTGCAATC 1259  
 QY 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 51  
 Db 1260 ACCAGGCTAACTGGCGGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTGCTCACT 1319  
 QY 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysThrPrpThrValTyrHisGly 71  
 Db 1320 GCTGCCAAACCTTCCTGGCAAGTGCATCAATGGGGTGTGTGGTGTCTTACCACGGG 1379

QY 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 91  
 Db 1380 GCCGAGCAGGAGCACCATCGCGTCACCCCAAGGTCTGTGTATCCAGATGTATACCAATGTA 1439  
 QY 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
 Db 1440 GACCAAGACCTTGTGGGTGGCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAC 1499  
 QY 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
 Db 1500 TCGCGCTCTCGGACCTTTACTGTGTACAGGACGCGGATGTCTATCCCGTGGCGCGG 1559  
 QY 132 ArgGlyAspSerArgLysSerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151  
 Db 1560 CGGGGTGATAGCAGGGCAGCTGTGTGCGCCCGGCCCATTTCTCTACTTCAAAGGCTCC 1619  
 QY 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171  
 Db 1620 TCGGGGGTCCGCTGTGTGCGCCCGGGGACGCGGTGGGCATATTTAGGGCGCGGTG 1679  
 QY 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
 Db 1680 TGCACCCGTGGAGTGGCTAAGCGGTGACTTTATCCCTGTGGAGAACCTAGACAAACC 1739  
 QY 192 MetArgSerPro 195  
 Db 1740 ATGAGGTCCCGC 1751  
 RESULT 14  
 AAN92106  
 ID AAN92106 standard; DNA; 7310 BP.  
 XX  
 AC AAN92106;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX  
 DE Combined open reading frames of the hepatitis C virus (HCV) cDNAs from  
 DE clones K9-1 through 15e.  
 XX  
 KW Hepatitis C virus; HCV; non-A, non-B hepatitis; NANBH.  
 XX  
 OS Hepatitis C virus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 3..7310  
 FT /\*tag= a  
 FT  
 XX  
 XX EP318216-A.  
 PN  
 XX 31-MAY-1989.  
 PD  
 XX 18-NOV-1988; 88EP-0310922.  
 PF  
 XX 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX Houghton M, Choo QL, Kuo G;  
 PI  
 XX WPI: 1989-159274/22.  
 DR P-PSDB; AAP2050.  
 XX  
 XX Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 XX  
 PS Claim 3; Figure 47-1 - 47-8; 139pp; English.  
 XX

CC It is a double-stranded nucleotide sequence of the open reading frame  
 CC (ORF) (tag a) extending through clones K9-1 to 15e of hepatitis C virus  
 CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to  
 CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)  
 CC it encodes could be used as immunoassay reagents and vaccines and to  
 CC generate antibodies useful in diagnosis and passive immunotherapy for  
 CC HCV infection/non-A, non-B hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;

Alignment Scores:  
 Pred. No.: 3.19e-71 Length: 7310  
 Score: 894.50 Matches: 175  
 Percent Similarity: 89.71% Conservative: 8  
 Best Local Similarity: 85.78% Mismatches: 10  
 Query Match: 87.61% Indels: 11  
 DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN92106 (1-7310)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
 Db 1665 CCCAGGGCCGGGAGATACGTCTCGGCCACGCCATGGATGGTCTCCAAAGGGGTGGAGG 1724  
 Qy 18 -----AlaTyAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31  
 Db 1725 TTGCTGGCGCCATCAGCGGTACGCCAGACAAAGGGCTCTCCTAGGGTGCATAATC 1784  
 Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnIleValSerThr 51  
 Db 1785 ACCAGCTTAACGTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTGCAACT 1844  
 Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 71  
 Db 1845 GCTGCCCAACCTCTCGGCAACGTGCATCAATGGGTGTCTGGACTGTCTACCACGGG 1904  
 Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 91  
 Db 1905 GCCGGAACGAGGACATCGCGTCAACCAGGGTCTCTCATCAGATGTATACCAATGTA 1964  
 Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
 Db 1965 GACCAAGACCTTGTGGCTGGCGCTGCCAGGTAGCGCTCATTTGACACCTTGCACT 2024  
 Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
 Db 2025 TCGCGCTCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCTCCGTGGCGCGG 2084  
 Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 151  
 Db 2085 CGGGGTGATAGCAGGGGCAGCGTCTCTCGCCCGGCCCATTTCTCTACTTGAAGGCTCC 2144  
 Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValAlaValGlyIlePheAlaAlaVal 171  
 Db 2145 TCGGGGGTTCGCTGTGTGCCCCCGGGGCACGCCGTGGGCATATTTAGGGCGCGGTG 2204  
 Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
 Db 2205 TGCACCCGTGAGTGGCTAAGCGGTGGACTTTATCTCTGTGGAGAACCTAGACACAAC 2264  
 Qy 192 MetArgSerPro 195  
 Db 2265 ATGAGGTCCCG 2276

RESULT 15

AAN90336

ID AAN90336 standard: DNA: 7310 BP.

XX

AC AAN90336;

XX

DT 25-MAR-2003 (updated)

DT 19-JUL-2001 (updated)  
 DT 01-NOV-1989 (first entry)  
 XX  
 DE Composite hepatitis C virus (HCV) cDNA.  
 KW Hepatitis C virus; cDNA; clone 15e; clone K9-1; probe; vaccine; ds.  
 XX Pan troglodytes.  
 XX GB2212511-A.  
 XX 26-JUL-1989.  
 XX 18-NOV-1988; 88GB-0027024.  
 XX 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 26-OCT-1988; 88US-0263584.  
 XX (CHIR ) CHIRON CORPORATION.  
 XX Houghton M. Choo QL, Kuo G;  
 XX WPI; 1989-215054/30.  
 DR P-PSDB; AAP90288.  
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment  
 PT of infection.  
 XX Disclosure; fig 47; 235pp; English.  
 XX The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by  
 CC aligning clones K9-1 through 15e in 5'-3' direction. The cDNA  
 CC encodes antigens which react with antibodies in patients with non-A  
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to  
 CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,  
 CC to raise antibodies for immunoassay or treatment, or to produce  
 CC vaccines. See also AAP90288, and AAN90303-35.  
 CC (N.B. This record was resubmitted to correct errors in the sequence.)  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.19e-71 Length: 7310  
 Score: 894.50 Matches: 175  
 Percent Similarity: 89.71% Conservative: 8  
 Best Local Similarity: 85.78% Mismatches: 10  
 Query Match: 87.61% Indels: 11  
 DB: 10 Gaps: 2

US-09-965-594-12 (1-195) x AAN90336 (1-7310)

Qy 3 LysLysGlySerValValIleValGly-----ArgIleValLeuAsnGly----- 17  
 Db 1665 CCCAGGGCCGGGAGATACGTCTCGGCCACGCCATGGATGGTCTCCAAAGGGGTGGAGG 1724  
 Qy 18 -----AlaTyAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 31  
 Db 1725 TTGCTGGCGCCATCAGCGGTACGCCAGACAAAGGGCTCTCCTAGGGTGCATAATC 1784  
 Qy 32 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGlnIleValSerThr 51  
 Db 1785 ACCAGCTTAACGTGGCGGGACAAACCAAGTGGAGGTGAGGTCCAGATTGTGCAACT 1844  
 Qy 52 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly 71  
 Db 1845 GCTGCCCAACCTCTCGGCAACGTGCATCAATGGGTGTCTGGACTGTCTACCACGGG 1904  
 Qy 72 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal 91  
 Db 1905 GCCGGAACGAGGACATCGCGTCAACCAGGGTCTCTCATCAGATGTATACCAATGTA 1964  
 Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111  
 Db 1965 GACCAAGACCTTGTGGCTGGCGCTGCCAGGTAGCGCTCATTTGACACCTTGCACT 2024  
 Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131  
 Db 2025 TCGCGCTCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCTCCGTGGCGCGG 2084  
 Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 151  
 Db 2085 CGGGGTGATAGCAGGGGCAGCGTCTCTCGCCCGGCCCATTTCTCTACTTGAAGGCTCC 2144  
 Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValAlaValGlyIlePheAlaAlaVal 171  
 Db 2145 TCGGGGGTTCGCTGTGTGCCCCCGGGGCACGCCGTGGGCATATTTAGGGCGCGGTG 2204  
 Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191  
 Db 2205 TGCACCCGTGAGTGGCTAAGCGGTGGACTTTATCTCTGTGGAGAACCTAGACACAAC 2264  
 Qy 192 MetArgSerPro 195  
 Db 2265 ATGAGGTCCCG 2276

```
Db 1905 GCCGGAACGAGGACCATCGCTACCCAAAGGTCTCTCATCCAGATGTATACCAATGTA 1964
Qy 92 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 111
   |||::|||
Db 1965 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGGTAGCCGCTCATTCACACCCCTGCAC 2024
Qy 112 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 131
   |||
Db 2025 TCGGGCTCTCGGACCTTACTGGTCACGAGGCACGCCGATGTCATTCCCGTGGCCGG 2084
Qy 132 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 151
   |||
Db 2085 CGGGGTGATAGCAGGGGAGCGCTGCTGCGCCCGGCCCATTTCTACTTGAAGGCTCC 2144
Qy 152 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 171
   |||
Db 2145 TCGGGGGGTCCGCTGTGTGCCCCGGGGGCACGCCGTGGGCATATTTAGGGCCCGGTG 2204
Qy 172 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 191
   |||
Db 2205 TGCACCCGTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 2264
Qy 192 MetArgSerPro 195
   |||
Db 2265 ATGAGGTCCCCG 2276
```

Search completed: August 30, 2003, 19:47:47  
Job time : 190.082 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1890.92 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-12

Perfect score: 1021

Sequence: 1 MKKGSWVIGRVLNGAYA.....VAKAVDFIPVESLETHMRSP 195

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09965594/runat\_29082003\_151919\_28322/app\_query.fasta\_1.2872  
-DB=EST -QPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.ecgn\_1\_12630\_@runat\_29082003\_151919\_28322 -NCPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estopl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: em\_gss\_hum: \*  
18: em\_gss\_inv: \*  
19: em\_gss\_pin: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_fun: \*  
22: em\_gss\_mam: \*  
23: em\_gss\_mus: \*  
24: em\_gss\_pro: \*  
25: em\_gss\_rod: \*  
26: em\_gss\_phg: \*  
27: em\_gss\_vrl: \*  
28: gb\_gss1: \*

29: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	106	10.4	984	10	BF304699
C 2	103.5	10.1	1199	13	BQ892487
C 3	103	10.1	814	13	BU148820
C 4	101.5	9.9	1403	13	BQ926101
C 5	100.5	9.8	844	12	BI198486
C 6	100.5	9.8	1204	13	BQ881847
C 7	99	9.7	615	12	BQ001625
C 8	99	9.7	643	12	BQ024121
C 9	99	9.7	754	12	BQ016176
C 10	98.5	9.6	961	10	BF203316
C 11	98	9.6	1141	11	AK080545
C 12	97.5	9.5	779	10	BF631437
C 13	97	9.5	1146	12	BM915803
C 14	96	9.4	701	10	BF863244
C 15	96	9.4	846	10	BF182274
C 16	95.5	9.4	901	10	BF307233
C 17	95.5	9.4	930	13	BU169585
C 18	95	9.3	407	9	AN785806
C 19	95	9.3	958	10	BG420860
C 20	95	9.3	1384	29	CC221189
C 21	95	9.3	1433	12	BM803824
C 22	94.5	9.3	641	9	AU127824
C 23	94.5	9.3	938	13	BQ894657
C 24	94	9.2	649	10	BE289911
C 25	94	9.2	940	14	CB993468
C 26	94	9.2	1283	13	BQ709745
C 27	93.5	9.2	701	14	CD262790
C 28	93.5	9.2	832	10	BG387051
C 29	93.5	9.2	905	13	BU542842
C 30	93.5	9.2	993	9	AL555424
C 31	93.5	9.2	1001	13	BQ928211
C 32	93	9.1	964	12	BI196460
C 33	92.5	9.1	556	14	CB216999
C 34	92.5	9.1	715	9	AU125614
C 35	92.5	9.1	846	13	BU540812
C 36	92.5	9.1	866	13	BX451426
C 37	92.5	9.1	881	14	CD105862
C 38	92.5	9.1	929	13	BQ672290
C 39	92.5	9.1	947	13	BU556872
C 40	92.5	9.1	979	13	BQ673186
C 41	92.5	9.1	1008	12	BI755608
C 42	92.5	9.1	1169	12	BM548430
C 43	92.5	9.1	1291	10	BE622016
C 44	92.5	9.1	1384	13	BQ919246
C 45	92	9.0	582	14	CB286751

# ALIGNMENTS

RESULT 1  
BF304699/c  
LOCUS 601888252F1 NIH\_MGC\_17 Homo sapiens CDNA clone IMAGE:4122276 5',  
DEFINITION mRNA sequence.  
ACCESSION BF304699  
VERSION BF304699.1 GI:11251586  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 984)

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LLC1005 row: g column: 13  
High quality sequence stop: 646.

**FEATURES**  
source Location/Qualifiers  
1..984  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:412276"  
/tissue\_type="rhandomysarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_17"  
/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 133 a 329 c 351 g 171 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.35 Length: 984  
Score: 106.00 Matches: 33  
Percent Similarity: 45.24% Conservative: 5  
Best Local Similarity: 39.29% Mismatches: 24  
Query Match: 10.38% Indels: 22  
DB: 10 Gaps: 5

US-09-965-594-12 (1-195) x BF304699 (1-984)

Qy	98	TrpProAlaProGlnGlySerArgSerLeuThr---	ProCysThrCysGlySerSerAsp	116
		:    :     :     :	:	
Db	646	TGSCCCAGTCCAGCGCATTCCTCGTGCGAAGAGACCGGTACCTGC-----	599	
Qy	117	LeuTyTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg	136	
		:     :     :     :	:	
Db	598	-----ACCAGGCACGACGACCAATACATCATCAAGGACAGCTGGT---	TCCCGC	554
Qy	137	GlySerLeuLeuSerProArgPro-----	IleSerTyTyrLeuLysGlySer	151
		:     :     :	:	
Db	553	GGCGCGCTCTGTGTGGGAAGACCTCGATGTGTCTCAAGCTCGCGCTCTGCTACTGTGAAGT	494	
Qy	152	SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal	171	
		:    :	:	
Db	493	CGCAGCTCCGTCAGTGCAGC-----	TTCCAGGCCCGCGGG	455
Qy	172	CysThrArgGly	175	
		:		
Db	454	TGGCGCCGAGGA	443	

**RESULT 2**  
**BQ892487**  
LOCUS BQ892487  
DEFINITION AGENCOURT\_8417538 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
IMAGE:6192708 5', mRNA sequence.  
ACCESSION BQ892487  
VERSION BQ892487.1 GI:22284501  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-f@mail.nih.gov  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13595 row: C column: 13  
High quality sequence start: 57  
High quality sequence stop: 394.  
Location/Qualifiers  
1. .ll199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6192708"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TGCAAGTCATGATGGCGAGCGGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupsaki, M.D./Ph.D. (aylor  
College of Medicine); available through Life  
Technologies."

BASE COUNT	255 a	362 c	343 g	211 t	28 others
ORIGIN					
Alignment Scores:					
Pred. No.:	9.74	Length:	1199		
Score:	103.50	Matches:	41		
Percent Similarity:	37.42%	Miscellaneous:	17		
Best Local Similarity:	26.45%	Mismatches:	53		
Query Match:	10.14%	Indels:	44		
DB:	13	Gaps:	6		
US-09-965-594-12 (1-195) x BQ892487 (1-1199)					
Qy	66	TrpThrValTyHisGlyAlaGlyThrArgThrTleAlaSerProLysGlyProValIle	85		
Db	484	TGGATCCATTTTAATAAAGGTGCTCTGTTAATCATGTGGCCACGGCCCCGCTGATA	543		
Qy	86	GlnMetTyThrAsnValAspLysAspLeuValGlyTrpProLaProlGInGlySerArg	105		
Db	544	CTTCCATTACCACATGTGACATGACTTT-----	573		
Qy	106	SerLeuThrProCysThr-----	CysGlySerSerAsp	116	
Db	574	---TGTCGTGCTGCACAGCACGCCCATGACCATGTGGGCTTATGTGAACGGCGAG	630		
Qy	117	LeuTyrr-LeuValThr-----	ArgHisAlaAspValIleProValArg	130	
Db	631	CGGTTTCATGGCCACTCCCTCCCTATATAAACACGCCAACGTCTTCATGGGCGGGCT	690		
Qy	131	-----	ArgArgGlyAspSerArgGlySerLeuLeu--	140	
Db	691	GGGTGTTTGCGACGGCGAACGGGGTGGGGGCATGGTAGGACTCGGGGGCGGATCTCTG	750		
Qy	141	-----SerProArgProfileSerTyrrLeuLys-----	GlySerSerGI	153	

```

751 AAAACCCACCTGGCCACCGATGCGCTAAGCCTCCCTTTACAAGCCACGCCCGCG 810
153 yGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaIleValCysTh 173
811 CCCCCCCTAACATCTCCCTACCCCTGCGCGCGGGGGGAGACGTGGCGCATACGGGC 870
173 rArgGlyValAlaLysAlaValAspPheIleProValGluSer 187
871 TCAGGGCGCTTTAAACGCCCGCGCCTTCGCCCGCGGCGGAAGCA 913

RESULT 3
BU148820
LOCUS BU148820 814 bp mRNA linear EST 03-SEP-2002
DEFINITION AGENCOURT_8742725 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6390808
5', mRNA sequence.
ACCESSION BU148820
VERSION BU148820.1 GI:22662352
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 814)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1378 row: a column: 17
High quality sequence stop: 381.
Location/Qualifiers
1. .814
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6390808"
/lab_host="NIH_MGC_129"
/note="Organ: olfactory epithelium; Vector:
pcmv-sport6.1.cdb; Site:1: EcoRV; Site:2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT 184 a 145 c 295 g 188 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.65 Length: 814
Score: 103.00 Matches: 45
Percent Similarity: 40.94% Conservative: 25
Best Local Similarity: 26.32% Mismatches: 63
Query Match: 10.09% Indels: 38
DB: 13 Gaps: 10

US-09-965-594-12 (1-195) x BU148820 (1-814)

QY 4 LysGlySerValValIleValGlyArgIleValLeuAsnGlyAlaThrAlaGlnThr 23
:::||||| |::: |::: |:::
126 AGAGGGAGTACTGCTGGAGGGGTACTATGCTGGAGGGGTACTGCTGGGAAAGGTA 185
:::||||| |::: |::: |:::
QY 24 ArgGlyGluGly---CysGlnGluThrSerGlnThrGlyArgAspLysAsnGlnVal 42
||| ||| ||| ||| |||
186 CTGTGCTGGAGGGGTACTGCTGAAGGAGT---ACTGCTGCTGAAGGATCTGCTG 242
||| ||| ||| ||| |||
QY 43 GluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsn 62
||| ||| ||| ||| |||

```

```

243 GAAGGTGATACTAGTTAGTGGAGAT-----|||
63 GlyVal---CysTrpThrValThrHisGlyAlaGlyThr-----ArgThrIleAla 78
||||| ||||| ||||| |||||
282 GGGTACTATGCTGAGGGGTACTATGCTGGAGGGGTACTGCTGGGAAAGGGTACTGT 341
||| ||||| ||||| |||||
79 SerProLysGlyProValIleGlnMetTyr-----ThrAsnValAspLys 93
::: ||||| ||||| ||||| |||||
342 GCTGGAGGGGTACTGCTGGAAAGGATACTGCTGAANGTGATATCTATTAGTGGAG 401
||| ||||| ||||| |||||
94 AspLeuVal-----GlyTrpProAlaProGlnGlySerArg 105
||| ||||| ||||| |||||
402 GATAGTCTGGAGTGGTACTGCTGGCGAGGATAGTGTCT-----CGAGGGAGA 455
||| ||||| ||||| |||||
106 SerLeuThrProCysThrCysGlySerSerAspLeuThrValThrArgHisAlaAsp 125
||||| ||||| ||||| |||||
456 TCTCTGGCTGCATGGCATGCTGTCACCTGCTGCTTCAT-----CACCGGAT 503
||| ||||| ||||| |||||
126 ValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle 145
||| ||||| ||||| |||||
504 -----TCTGGTCTGAGCTGGCTGAAATATCTCTAAACCCGCT 545
||| ||||| ||||| |||||
146 SerTyrLeuLysGlySerSerGlyGlyProLeu 156
::: ||||| ||||| |||||
546 TCTTTTCCAGGCTGCTGACGCTTTTCCCGCT 578
||| ||||| ||||| |||||

RESULT 4
BU226101
LOCUS BU226101 1403 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION BU226101
VERSION BU226101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1378 row: j column: 07
High quality sequence stop: 101.
Location/Qualifiers
1. .1403
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="NIH_MGC_130"
/note="Organ: oocytes; Vector: pcmv-sport6.1.cdb;
Site:1: EcoRV; Site:2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 297 a 521 c 237 g 345 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 18.5 Length: 1403

```



```

Db      195 CCGGAGGACATCGGATCTCTCTGATCTGCTTGGAGCCACATGTCTGCTCC 139
RESULT 6
BQ881847/c
LOCUS   BQ881847 1204 bp mRNA linear EST 16-AUG-2002
DEFINITION AGENCOURT_8712410 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6295246
5', mRNA sequence.
ACCESSION BQ881847
VERSION   BQ881847.1 GI:22273855
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1204)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLES    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cyapbs-r@mail.nih.gov
          Tissue Procurement: DCTD/DTF
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LHC2501 row: c column: 23
          High quality sequence stop: 322.
          Location/Qualifiers
            1..1204
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6295246"
              /tissue_type="melanotic melanoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_112"
              /notes="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGCACAGAG(G). Library constructed by Ling Hong in the
              laboratory of Gerald M. Rubin (University of California,
              Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
              Superscript II RT (Life Technologies). Note: this is a
              NIH_MGC Library."
BASE COUNT 228 a 430 c 265 g 280 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 19 Length: 1204
Score: 100.50 Matches: 41
Percent Similarity: 36.02% Conservative: 17
Best Local Similarity: 25.47% Mismatches: 58
Query Match: 9.84% Indels: 45
DB: 13 Gaps: 6
US-09-965-594-12 (1-195) x BQ881847 (1-1204)
Qy      10 ValGlyArgLeuValLeuAsnGlyAlaTyrAlaGlnGlnThrArgGly----- 25
Db      773 GTGCCAAGGGCTGTGTCACGGGGCAGTATATGCGCCAAACCGTGGTGGCAACGCC 714
Qy      26 -----GluGluGlyCysGlnGlnThrSerGlnThrGlnThrGlyArgAspLysAsn 40
Db      713 GGATAAAGGGCGCAACAGCGGGCGGCTGGCTAATTAAGAGGAGGTGTAACAAC 654
Qy      41 GlnValGluGlyGluValGlnLeuValSerThrAlaAlaGlnThrPheLeuAlaThrCys 60
Db      653 CGCGGGGAGGGCGCAAGGAATATATCTGGAGGTGTCCGGGGGGCCATAGGGAGGGA 594
Qy      61 IleAsnGlyValCysTrp-----ThrValTyrHisGlyAlaGlyThrArgThrIle 77
          ::::: |||||

```

```

Db      593 CACGACGGGAAGATGTTGGTTCAGGATGCGGGGAAAGAGCGGGGGGAGGCTATA 534
Qy      78 AlaSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGly 97
          ||| ::::: |||||
Db      533 GGGGGGGAGGAGGTGGGTTAATT-----ATAGTGGGT 501
Qy      98 -----TTPProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySer 114
          ||||| ||::: |||||
Db      500 GTATAACGGTGGCGGGGCTCGTGGTTCTCG----- 468
Qy      115 SerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp 134
          ::::: ||||| ::::: |||
Db      467 -----GTCTCCCGACCTTGTGGGATTGTGGGTTACCCGGGTCTGGGAG 423
Qy      135 SerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGly 154
          ::::: ||| ::::: |||||
Db      422 ACGGAGGGGGGATGTATCG-----AAGGAGGGGGGGGGA 384
Qy      155 Pro 155
          |||
Db      383 CCG 381
          |||
RESULT 7
BQ001625/c
LOCUS   BQ001625 615 bp mRNA linear EST 05-DEC-2001
DEFINITION BQ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
mRNA sequence.
ACCESSION BQ001625
VERSION   BQ001625.1 GI:17364516
KEYWORDS  EST.
SOURCE    Oryzias latipes (Japanese medaka)
ORGANISM  Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 615)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Jinbo,T. and Takeda,H.
TITLES    Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT    Contact: Tadasu Shin-i
          Center for Genetic Resource Information
          National Institute of Genetics
          1111 Yata, Mishima, Shizuoka 411-8540, Japan
          Tel: 81-559-81-6856
          Fax: 81-559-81-6855
          Email: tshiniegenes.nig.ac.jp.
FEATURES
          Location/Qualifiers
            1..615
              /organism="Oryzias latipes"
              /mol_type="mRNA"
              /strain="Hd-r"
              /db_xref="taxon:8090"
              /clone="MF01SSA025C02"
              /sex="mixture of female and male"
              /tissue_type="whole embryo"
              /dev_stage="segmentation stage 20 - 25"
              /clone_lib="MF01SSA cDNA"
BASE COUNT 140 a 166 c 165 g 144 t
ORIGIN
Alignment Scores:
Pred. No.: 11.3 Length: 615
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.70% Indels: 50
DB: 12 Gaps: 7
US-09-965-594-12 (1-195) x BQ001625 (1-615)
Qy      39 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
          ||||| ||||| :::::

```



```

Db      511 AAAATGACGTAGAACCAACACACACAGTCCACACACATGTTCTGTTCTACGGGCT 452
QY      59  ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78
Db      451 -----TGTGGAGAACCTATCACAGTTCCTGCTTTAGACCAACGGCA 410
QY      79  SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 93
Db      409 GCTCTCGCGCGGAGAGCTCTCTGGCCAGTTGTG----- 374
QY      94  AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 109
Db      373 -----ACTGTGGAGACCAAGAGCGTCACCCGGAGCTGTAGG 335
QY      110 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 122
Db      334 CTGACGGATGCGGATGGCTCTGCT-----TTGGTTCTCTGCTCTCTGATCA 284
QY      123 -----HisAlaAspValIleProValArgArgArgGlyAspSer 135
Db      283 TCTTCTCCTACCTGACCTTCCACATCCAGGTGTCGCCAGCGTGGTCTGACGGGTGATGG 224
QY      136 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySer 152
Db      223 AGAGCGCGACAGCGACGTGGGGTGAATCTCTGCAGGACGCTTCTCAGCGCGGATCA 164
QY      153 GlyGlyProLeuLeuCysProAlaGlyHisAla 163
Db      163 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 131

RESULT 8
BJ024121 643 bp mRNA linear EST 05-DEC-2001
LOCUS    BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',
DEFINITION mRNA sequence.
ACCESSION BJ024121
VERSION    BJ024121.1 GI:17377389
KEYWORDS
SOURCE
ORGANISM  Oryzias latipes (Japanese medaka)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 643)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE      Medaka EST Project in Takeda's lab
JOURNAL    Unpublished
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
FEATURES   source
            1..643
            /organism="Oryzias latipes"
            /mol_type="mRNA"
            /strain="Hd-r"
            /db_xref="taxon:8090"
            /clone="MF01SSA143D12"
            /sex="mixture of female and male"
            /tissue_type="whole embryo"
            /dev_stage="segmentation stage 20 - 25"
            /clone_lib="MF01SSA cDNA"
BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN
Alignment Scores:
Pred. No.: 12 Length: 643
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9

```

```

Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.70% Indels: 50
DB: 12 Gaps: 7
US-09-965-594-12 (1-195) x BJ024121 (1-643)

QY      39  LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 58
Db      242 AAAATGACGTAGAACCAACACACACATGTTCTGTTCTACGGGCT 301
QY      59  ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 78
Db      302 -----TGTGGAGAACCTATCACAGTTCCTGCTTTAGACCAACGGCA 343
QY      79  SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 93
Db      344 GCTCTCGCGCGGAGAGCTCTCTGGCCAGTTGTG----- 379
QY      94  AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 109
Db      380 -----ACTGTGGAGACCAAGAGCGTCACCCGGAGCTGTAGG 418
QY      110 -----CysThrCysGlySerAspLeuTyrLeuValThrArg----- 122
Db      419 CTGACGGATGCGGATGGCTCTGCT-----TTGGTTCTCTGCTCTCTGATCA 469
QY      123 -----HisAlaAspValIleProValArgArgGlyAspSer 135
Db      470 TCTTCTCCTACCTGACCTTCCACATCCAGGTGTCGCCAGCGTGGTCTGACGGGTGATGG 529
QY      136 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySer 152
Db      530 AGAGCGCGACAGCGACGTGGGGTGAATCTCTGCAGGACGCTTCTCAGCGCGGATCA 589
QY      153 GlyGlyProLeuLeuCysProAlaGlyHisAla 163
Db      590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 9
BJ016176 754 bp mRNA linear EST 05-DEC-2001
LOCUS    BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
DEFINITION mRNA sequence.
ACCESSION BJ016176
VERSION    BJ016176.1 GI:17376695
KEYWORDS
SOURCE
ORGANISM  Oryzias latipes (Japanese medaka)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 754)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE      Medaka EST Project in Takeda's lab
JOURNAL    Unpublished
COMMENT    Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
FEATURES   source
            1..754
            /organism="Oryzias latipes"
            /mol_type="mRNA"
            /strain="Hd-r"
            /db_xref="taxon:8090"
            /clone="MF01SSA025C02"
            /sex="mixture of female and male"
            /tissue_type="whole embryo"
            /dev_stage="segmentation stage 20 - 25"

```







```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482056"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      169 a   492 c   344 g   141 t
ORIGIN

Alignment Scores:
Pred. No.:      38.9      Length:      1146
Score:          97.00     Matches:      49
Percent Similarity: 34.05%   Conservative: 14
Best Local Similarity: 26.49% Mismatches:      65
Query Match:      9.50%     Indels:       57
DB:              12        Gaps:         10

US-09-965-594-12 (1-195) x BM915803 (1-1146)
QY      17 GlyAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSer-GlnThrG1 36
Db      1107 GGAGGGCGGGCGAGAGGGTGTGGCGAGGAGGGTGGCTCTCCGGCTCTCGTACCGT 1048
QY      36 yArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPh 56
Db      1047 CCGTGGGAGCTGGAGGGACAGAGCTACGCGCTGGGTAGGGACGCGCT----- 996
QY      56 eLeuAlaThrCysIleAsnGlyValCysTyrThrValThrHis-----GlyAl 72
Db      995 -----GGTGGGATGTTGTGG-----TATCACTCCCGCGCGGGGGG 958
QY      72 aGlyThr-----ArgThrIleAlaSerPro-----LysGlyProValI1 85
Db      957 AGTTACGTGACGGAGGGCGCGGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
QY      85 eGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGln----- 102
Db      902 -----CAGATGTGGGGTGGGAAGCGCGCGCTCGCGCGGT 868
QY      103 -GlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrAr 122
Db      867 GGGGCCACAGACTGCTCTCGTGTCTTCTGTGG----- 834
QY      122 gHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerPr 142
Db      833 -----CGGAGGGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 796
QY      142 oArgProIleSerTyrLeuLysGlySerSerGlyClyProLeuLeuCysProAlaGlyHi 162
Db      795 TCCCGCTCTCGGTATCTACAGGCGCGCGAGGAGGACCATACCTCTCTCCG----- 744
QY      162 sAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPh 182
Db      743 -----TGGCGCTTCCCGGCTGTGTGTCTCGCGGTCTCGCGGGGGGGGGGGT 691
QY      182 e----IleProVal 185
Db      690 TCGCGTACCTTG 678

RESULT 14
LOCUS      BF863244
DEFINITION 963042C02.x1 C. reinhardtii CC-1690, Stress condition I, normalized
           , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

```

```

ACCESSION      BF863244
VERSION        BF863244.1
KEYWORDS       GI:12253388
SOURCE         Chlamydomonas reinhardtii
ORGANISM       Chlamydomonas reinhardtii
REFERENCE      1 (bases 1 to 701)
AUTHORS        Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
                Lefebvre, P., McDermott, J.P., Shrager, J., Sliflow, C. and Stern, D.
TITLE          Analyses of the Chlamydomonas reinhardtii Genome: A Model,
                Unicellular System for Analyzing Gene Function and Regulation in
                Vascular Plants; project phase 3
JOURNAL        Unpublished
COMMENT        Contact: Charles Hauser
                DCMB Box 91000
                Duke University
                Durham, NC 27708-1000
                Tel: 919 613 8159
                Fax: 919 613 8177
                Email: chauser@duke.edu.
FEATURES       source
                1..701
                /organism="Chlamydomonas reinhardtii"
                /mol_type="mRNA"
                /strain="CC-1690 wild type mt+ 21gr"
                /db_xref="taxon:3055"
                /clone_lib="C. reinhardtii CC-1690, Stress condition I,
                normalized, Lambda Zap II"
                /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                XhoI; This library, constructed by John Davies and Jeffrey
                McDermott, combines cDNAs from CC-1690 cells grown to
                mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
                1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
                4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
                purified from each sample, pooled and cDNA synthesized.
                The cDNA was directionally cloned into lambda Zap II
                (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                pBluescript II SK- plasmids were excised from the lambda
                Zap clones by superinfection with ExAssist (Stratagene)
                phage. The library was normalized using method 4 described
                in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT     173 a   213 c   175 g   140 t
ORIGIN

Alignment Scores:
Pred. No.:      26      Length:      701
Score:          96.00     Matches:      32
Percent Similarity: 40.71%   Conservative: 14
Best Local Similarity: 28.32% Mismatches:      45
Query Match:      9.40%     Indels:       22
DB:              10        Gaps:         4

US-09-965-594-12 (1-195) x BF863244 (1-701)
QY      69 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 84
Db      171 CACCACCATACCTTGTCTCTCAGCTTCACACACCAAAATATGCCATACGGCCACTA 230
QY      85 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGlnGlySer 104
Db      231 ACAAGTTATCATACAGG-----AAGGACACCGCGCTTGGCCACCCCTTGGAGCGG 284
QY      105 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 124
Db      285 AGAAGCCGACCGCTGGTCTCTGGGGTATCCGCATGCTATGCATCTCCCGCTATCAG 344
QY      125 AspValIle-----ProValArgArgGlyAspSerArg----- 136
Db      345 GAGATCATTTGCGTGTGGCTTTAGTCACCCCAAGAGAGCGCTGGGATGGGCATTTATAA 404
QY      137 -----GlySerLeuLeuSerProArgProIleSerTyrLeu 148

```

```

Db      405 GAAGGGACGGCAATTCCTGTCGGGAAAAGTCAGGCCCCCAAGGTCGTGACCAAGTGCTA 464
QY      149 LysGlySerSerGlyGlyProLeuLeuCysProAlaGly 161
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      465 CTCRAAGGACGCAATGGGAGCCTTTTCGGCGGTGCGGGT 503

RESULT 15
BF182274/c
LOCUS   BF182274
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE  BF182274.1 GI:11060416
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 846)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cyapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9308 row: g column: 07
High quality sequence stop: 696.
FEATURES
      source
      1..846
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="IMAGE:4035102"
      /tissue_type="tumor, gross tissue"
      /dev_stage="7 months"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Mam5"
      /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
      Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
      Library constructed by Life Technologies. Investigators
      providing samples: Lothar Hennighausen/Robin Humphreys,
      NIH"
BASE COUNT      176 a      218 c      241 g      210 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      33      Length:      846
Score:          96.00      Matches:      46
Percent Similarity: 47.20%      Conservative: 13
Best Local Similarity: 36.80%      Mismatches: 36
Query Match:      9.40%      Indels:      31
DB:              10      Gaps:      9

US-09-965-594-12 (1-195) x BF182274 (1-846)
QY      71 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAs 90
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      757 GGTTCCTCTACCAACAGCCCTGGATGAGAAGGACCA-----CATCCTTC 710
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      90 nValAspLysAspLeuValGlyTrp-----ProAlaProGl 102
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      709 GGTTCCTTCAGTCCCAAGCTGGAGAAAGTGAACACAGAGAAAGAGGACGGTCCCTCA 650
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      102 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVa 120
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      649 GTCCCCACGCTTCAGTAGTCTAGACAAAGTGTCTGTCTGGA----- 610

```

```

QY      120 lThrArgHisAlaAspValIleProValIleArgArgGlyAspSerArgGlySerLeuLe 140
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      609 -ACTAGACACACCT--GTAATCCAGGAGGAGCGNTGGAGGAACAGAGGACTCC---CT 556
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      140 lUserProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu---LeuCysPr 159
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      555 GACCCCACTCC---TCCGTCTAGCGGACCTCTCTGCCCCCACTCCCTCTGTGTCC 499
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      159 cAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 174
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      498 TAGTGGGACACCTCTCCCGAGGACACAGCACTGTACTCCCTTTGGCCCTCTGCACTCT 439
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY      174 gGlyValAlaLys 178
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db      438 TGGGATGACTGAG 426
      111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

Search completed: August 31, 2003, 04:27:24  
Job time : 1896.92 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKGSVVIGRINLSGDTA.....VAKAVDPVSELETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1032	100.0	197	21	AA15221 Hepatitis C virus
2	1015	98.4	197	21	AA15222 Hepatitis C virus
3	998	96.7	195	21	AA15220 Hepatitis C virus
4	993	96.2	197	21	AA15226 Hepatitis C virus
5	985	95.4	197	21	AA15223 Hepatitis C virus
6	973	94.3	197	21	AA15224 Hepatitis C virus
7	963	93.3	197	21	AA15225 Hepatitis C virus
8	959	92.9	195	21	AA15212 Hepatitis C virus
9	902.5	87.5	3011	14	AA40120 HCV genomic amino

10	901.5	87.4	1766	10	AA192041	Sequence encoded 1
11	901.5	87.4	1786	10	AA190158	Protein sequence 0
12	901.5	87.4	2261	10	AA190164	Peptide encoded by
13	901.5	87.4	2301	10	AA192047	Sequence encoded 1
14	901.5	87.4	2436	10	AA192050	Sequence encoded 1
15	901.5	87.4	2436	10	AA190288	Peptide encoded by
16	901.5	87.4	2772	21	AA18540	Protein encoded by
17	901.5	87.4	2816	14	AA184009	HCV-1 polyprotein.
18	901.5	87.4	2894	16	AA170230	Composite hepatiti
19	901.5	87.4	2955	20	AA14975	Amino acid sequenc
20	901.5	87.4	2955	21	AA18541	Polyprotein encode
21	901.5	87.4	3011	13	AA121519	Compiled HCV sequ
22	901.5	87.4	3011	14	AA131621	Hepatitis C virus
23	901.5	87.4	3011	17	AA190931	Hepatitis C virus
24	901.5	87.4	3011	18	AA134480	HCV polyprotein.
25	901.5	87.4	3011	19	AA140038	HCV polyprotein.
26	901.5	87.4	3011	23	AA122049	Hepatitis C virus
27	901.5	87.4	3011	23	AA184597	HCV polyprotein la
28	900.5	87.3	3011	15	AA186995	Hepatitis C virus
29	899	87.1	182	21	AA15211	Hepatitis C virus
30	899	87.1	609	15	AA151170	Hepatitis C virus
31	899	87.1	631	18	AA131884	A nonstructural pr
32	899	87.1	686	23	AA18689	HCV-1 NS3/4a mutan
33	899	87.1	686	23	AA176377	Hepatitis C virus
34	899	87.1	686	24	AB172261	Hepatitis C virus
35	898.5	87.1	2435	13	AA125135	HCV-1 NS3/4a confo
36	898.5	87.1	2436	13	AA128582	HCV amino acid seq
37	897.5	87.0	2772	11	AA108123	Hepatitis C virus
38	897	86.9	631	20	AA193482	HCV NS3 protein.
39	896.5	86.9	3011	19	AA177397	Hepatitis C virus
40	896.5	86.9	3011	24	AB171460	Amino acid sequenc
41	896.5	86.9	3012	23	AA109289	Hepatitis C virus
42	895	86.7	632	23	AA121847	Hepatitis C virus
43	895	86.7	632	23	AA119005	Hepatitis C virus
44	895	86.7	686	23	AA121837	Hepatitis C virus
45	895	86.7	686	23	AA121838	Hepatitis C virus

#### ALIGNMENTS

RESULT 1  
AA15221  
ID AA15221 standard; protein; 197 AA.  
XX  
AA15221;  
XX  
19-DEC-2000 (first entry)  
XX  
Hepatitis C virus NS4A-NS3 fusion protease #3.  
DE  
Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW  
liver failure; liver cancer; mutant; muteln.  
XX  
Hepatitis C virus.  
OS  
Synthetic.  
XX  
WO200040707-A1.  
XX  
13-JUL-2000.  
XX  
06-JAN-2000; 2000WO-0500345.  
XX  
08-JAN-1999; 99US-0115271.  
XX  
(BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
WPI; 2000-465976/40.  
DR  
N-PSDB; AAA73330.  
XX  
Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
XX  
PS Claim 23; Fig 13; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
CC variant.  
XX  
XX  
SQ Sequence 197 AA;  
Query Match 100.0%; Score 1032; DB 21; Length 197;  
Best Local Similarity 100.0%; Pred. No. 1.1e-98;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60  
DB 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60  
QY 61 TCINGVCWTVYHGAGTRTIA SPKGPVIOMYTNVDKLVGWPAQGSRLTCTCGSSDLY 120  
DB 61 TCINGVCWTVYHGAGTRTIA SPKGPVIOMYTNVDKLVGWPAQGSRLTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 2  
AAB15222  
ID AAB15222 standard; protein: 197 AA.  
XX  
XX AAB15222;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #4.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
XX 06-JAN-2000; 2000WO-US00345.  
PF  
XX 08-JAN-1999; 99US-0115271.  
PR  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX WittekInd M, Weinheimer S, Zhang Y, Goldfarb V;  
PI  
XX WPI: 2000-465976/40.  
DR  
XX N-PSDB; AAA73331.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
XX  
PS Claim 23; Fig 14; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
CC variant.  
XX  
XX  
SQ Sequence 197 AA;  
Query Match 98.48; Score 1015; DB 21; Length 197;  
Best Local Similarity 98.5%; Pred. No. 6.2e-97;  
Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60  
DB 1 MKKKGSVIVGRINLSGDTAYAAQTGRGEGCGQETSGTGRDNQVEGEVQIVSTAAQTFLA 60  
QY 61 TCINGVCWTVYHGAGTRTIA SPKGPVIOMYTNVDKLVGWPAQGSRLTCTCGSSDLY 120  
DB 61 TCINGVCWTVYHGAGTRTIA SPKGPVIOMYTNVDKLVGWPAQGSRLTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRRGDSRGLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 3  
AAB15220  
ID AAB15220 standard; protein: 195 AA.  
XX  
XX AAB15220;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
XX 06-JAN-2000; 2000WO-US00345.  
PF  
XX 08-JAN-1999; 99US-0115271.  
PR  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX WittekInd M, Weinheimer S, Zhang Y, Goldfarb V;  
PI  
XX WPI: 2000-465976/40.  
DR  
XX N-PSDB; AAA73329.



XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT  
XX  
PS Claim 23: Fig 12: 66pp: English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
CC variant.  
XX  
SQ Sequence 195 AA;  
  
Query Match 96.7%; Score 998; DB 21; Length 195;  
Best Local Similarity 98.0%; Pred. No. 3.5e-95;  
Matches 193; Conservative 1; Mismatches 1; Indels 2; Gaps 1;  
  
QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCOFTSOTGRDKNOVEGEVQIVSTAAQTFLA 60  
DB 1 MKKKGSVIVGRIVLNG--AYAAQTRGEECCOFTSOTGRDKNOVEGEVQIVSTAAQTFLA 58  
  
QY 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLTPTCTCGSSDLY 120  
DB 59 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLTPTCTCGSSDLY 118  
  
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPGLCPAGHAGVIFRAAVCTRGVAK 180  
DB 119 LVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPGLCPAGHAGVIFRAAVCTRGVAK 178  
  
QY 181 AVDFIPVESLETTMRSP 197  
DB 179 AVDFIPVESLETTMRSP 195  
  
RESULT 4  
AAB15226  
ID AAB15226 standard; protein; 197 AA.  
XX  
AC AAB15226;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #8.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI; 2000-465976/40.

DR N-PSDB; AAA73335.  
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT  
XX  
PS Example 5; Fig 18: 66pp: English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0  
CC wild-type sequence.  
XX  
SQ Sequence 197 AA;  
  
Query Match 96.2%; Score 993; DB 21; Length 197;  
Best Local Similarity 97.5%; Pred. No. 1.2e-94;  
Matches 192; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGECCOFTSOTGRDKNOVEGEVQIVSTAAQTFLA 60  
DB 1 MKKKGSVIVGRINLSGDTAYAAQOTRGLGCIITSUTGRDKNOVEGEVQIVSTAAQTFLA 60  
  
QY 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLTPTCTCGSSDLY 120  
DB 61 TCINGVCWTYYHGAGTRTIIASPKGPVIQMYTNVDKDLVGPAPQGSRSLSLTPTCTCGSSDLY 120  
  
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPGLCPAGHAGVIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGSSGGPGLCPAGHAGVIFRAAVCTRGVAK 180  
  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
  
RESULT 5  
AAB15223  
ID AAB15223 standard; protein; 197 AA.  
XX  
AC AAB15223;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #5.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI; 2000-465976/40.

DR WPI: 2000-465976/40.  
 XX N-PSDB; AAA73332.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX

PS Claim 23; Fig 15; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX

SQ Sequence 197 AA;

Query Match 95.4%; Score 985; DB 21; Length 197;  
 Best Local Similarity 97.0%; Pred. No. 8e-94;  
 Matches 191; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKKGSVIVGRINLSGDTAYAAQTGRGEGCQETSQTGRDNQVEGEVQIVSTAAQTFLA 60  
 DB 1 MKKGSVIVGRINLSGDTAYAAQTGRGEGCQETSQTGRDNQVEGEVQIVSTAAQTFLA 60  
 QY 61 TCINGVCTVYHAGCTRTIASPKGPVIOYTNVDKDLVGPAPQGSRLTPTCGSSDLY 120  
 DB 61 TSINGVLTVYHAGCTRTIASPKGPVIOYTNVDKDLVGPAPQGSRLTPTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAVCTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAVCTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 6

AAB15224  
 ID AAB15224 standard; protein; 197 AA.

XX AAB15224;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #6.

XX hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.  
 OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX

DR WPI: 2000-465976/40.

DR N-PSDB; AAA73333.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX

PS Claim 23; Fig 16; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX

SQ Sequence 197 AA;

Query Match 94.3%; Score 973; DB 21; Length 197;  
 Best Local Similarity 95.4%; Pred. No. 1.4e-92;  
 Matches 188; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKKGSVIVGRINLSGDTAYAAQTGRGEGCQETSQTGRDNQVEGEVQIVSTAAQTFLA 60  
 DB 1 MKKGSVIVGRINLSGDTAYAAQTGRGEGCQETSQTGRDNQVEGEVQIVSTAAQTFLA 60  
 QY 61 TCINGVCTVYHAGCTRTIASPKGPVIOYTNVDKDLVGPAPQGSRLTPTCGSSDLY 120  
 DB 61 TSINGVLTVYHAGCTRTIASPKGPVIOYTNVDKDLVGPAPQGSRLTPTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAVCTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCAGHAGVIFRAAVCTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 7

AAB15225  
 ID AAB15225 standard; protein; 197 AA.

XX AAB15225;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #7.

XX hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.

XX Hepatitis C virus.  
 OS Synthetic.

XX WO2000040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI: 2000-465976/40.  
DR N-PSDB: AAR73334.  
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
XX Claim 23; Fig 17: 66pp; English.  
PS  
XX The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7  
XX variant.  
XX  
SQ Sequence 197 AA;  
Query Match 93.3%; Score 963; DB 21; Length 197;  
Best Local Similarity 94.9%; Pred. No. 1.5e-91;  
Matches 187; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSQTGRDKNOVEGEVQIVSTAQTFLA 60  
DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQGTQKTSHTGRDKNOVEGEVQIVSTAQTFLA 60  
QY 61 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120  
DB 61 TSINGVLTATVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 8  
AAB15212  
ID AAB15212 standard; protein; 195 AA.  
XX  
AC AAB15212;  
XX  
XX 19-DEC-2000 (first entry)  
DT  
DE Hepatitis C virus NS4A-NS3 fusion protease #1.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
XX WO200040707-A1.  
PN  
XX  
XX 13-JUL-2000.  
PD  
XX  
XX 06-JAN-2000; 2000WO-US00345.  
PF  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
PI WPI: 2000-465976/40.  
DR N-PSDB: AAR73328.  
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
XX Example 2; Fig 10: 66pp; English.  
PS  
XX The present sequence is a fusion protein created using the Hepatitis C  
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
CC essential for the replication of the virus, acting to cleave its  
CC replicative proteins from the polyprotein produced from the HCV genome.  
CC Inhibitors of the two proteins should be effective as antiviral  
CC treatments of HCV infection. This is useful as HCV can lead to chronic  
CC liver disease such as cirrhosis, liver failure and liver cancer. The  
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
CC proteins which can be used to identify inhibitors of this type, as well  
CC as enabling structural studies of the protease and protease-inhibitor  
CC complexes.  
XX  
SQ Sequence 195 AA;  
Query Match 92.9%; Score 959; DB 21; Length 195;  
Best Local Similarity 95.4%; Pred. No. 3.9e-91;  
Matches 188; Conservative 1; Mismatches 6; Indels 2; Gaps 1;  
QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSQTGRDKNOVEGEVQIVSTAQTFLA 60  
DB 1 MKKGSVVIVGRIVLNG--AYAQOTRGLGCIITSLTGRDKNOVEGEVQIVSTAQTFLA 58  
QY 61 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 120  
DB 59 TCINGVCTVYHGAGTRTIAASPKGPVQMTNVNVDKDLVGVWPAQGSRSLSLTCTCGSSDLY 118  
QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAK 180  
DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPLLCPAGHAVGIFRAAVCTRGVAK 178  
QY 181 AVDFIPVESLETTMRSP 197  
DB 179 AVDFIPVESLETTMRSP 195  
RESULT 9  
AAR40120  
ID AAR40120 standard; protein; 3011 AA.  
XX  
AC AAR40120;  
XX  
XX 25-MAR-2003 (updated)  
DT 27-JAN-1994 (first entry)  
DE  
DE HCV genomic amino acid sequence isolated from infected human LG.  
XX  
KW Hepatitis C Virus; Non-A, non-B hepatitis virus; HCV; NANBHV;  
KW human growth hormone; HGH; secretion signal; fusion protein;  
KW vaccine.  
XX  
OS Hepatitis C Virus.  
XX  
XX WO9315193-A1.  
PN  
XX  
XX 05-AUG-1993.  
PD  
XX  
XX 29-JAN-1993; 93WO-US00907.  
PF  
XX  
XX 31-JAN-1992; 92US-0830024.  
PR  
XX

PA (ABBO ) ABBOTT LAB.  
 XX Bode SL, Casey JM, Desai SM, Devare SG, Frail DE;  
 PI Yamaguchi J, Zeck BJ;  
 XX WPI; 1993-258673/32.  
 XX New plasmid pHCV-162 is a mammalian expression systems for HCV  
 PT proteins - useful for diagnosing HCV infection and as vaccines  
 PT for preventing HCV infection  
 XX  
 XX Example 1; Page 39-49; 100pp; English.  
 PS  
 CC RNA was isolated from the plasma of a HCV seropositive human  
 CC (designated "1G") and cDNA was prepared from it. The cDNA was  
 CC PCR amplified using specific primers with sequences based  
 CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further  
 CC amplification using nested primers resulted in 7 adjacent HCV DNA  
 CC fragments which could be assembled into a full-length sequence. The  
 CC DNA sequence was determined and translated into the genomic amino  
 CC acid sequence. Comparison of the 1G genomic amino acid sequence  
 CC with that from HCV-1 showed 134 amino acid differences.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 3011 AA;

Query Match 87.5%; Score 902.5; DB 14; Length 3011;  
 Best Local Similarity 85.8%; Pred. No. 1.1e-83;  
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;  
 QY 3 KKGSVIVGRIN-----LSGDTAYAOQTREEGCQETSOTGRDKNOVEGEVQIVST 53  
 DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAOQTRGLLGCIIITSLTGRDKNOVEGEVQIVST 1064  
 QY 54 AAQTFLATCINGCVTVYHGAGTRTIA SPKGPVIOMTYNDKDLVGPAPGQSRSLTPCT 113  
 DB 1065 AAQTFLATCINGCVTVYHGAGTRTIA SPKGPVIOMTYNDKDLVGPAPGQSRSLTPCT 1124  
 QY 114 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRISYLYKSGSGGLLCPAGHAVGIFRAAV 173  
 DB 1125 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRISYLYKSGSGGLLCPAGHAVGIFRAAV 1184  
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 10  
 AAP92041  
 ID AAP92041 standard; protein; 1766 AA.  
 XX  
 AC AAP92041;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX  
 XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones  
 DE 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f,  
 DE 33f, 33g and 39c.  
 XX  
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)  
 KW  
 OS Hepatitis C virus.  
 XX  
 XX EP318216-A.  
 XX  
 XX 31-MAY-1989.  
 PD  
 XX 18-NOV-1988; 88EP-0310922.  
 PF  
 XX 18-NOV-1987; 87US-0122714.  
 XX 30-DEC-1987; 87US-0139886.  
 PR  
 XX 26-FEB-1988; 88US-0161072.  
 PR

PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Houghton M, Choo QL, Kuo G;  
 PI  
 XX WPI; 1989-159274/22.  
 DR N-PSDB; AAN92097.  
 XX  
 PT Purified hepatitis C virus  
 PT - and associated nucleic acids and polypeptide(s)  
 XX  
 XX Claim 13; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.  
 PS  
 CC It is the sequence encoded in the open reading frame of hepatitis C virus  
 CC cDNA inserts in clones 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36,  
 CC 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could  
 CC be used in immunoassay reagents and vaccines and to generate antibodies  
 CC useful in diagnosis and passive immunotherapy for HCV infection/non-A.  
 CC non-B hepatitis.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 XX  
 SQ Sequence 1766 AA;

Query Match 87.4%; Score 901.5; DB 10; Length 1766;  
 Best Local Similarity 85.8%; Pred. No. 6.9e-84;  
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;  
 QY 3 KKGSVIVGRIN-----LSGDTAYAOQTREEGCQETSOTGRDKNOVEGEVQIVST 53  
 DB 289 RRGREILLGPADGMVSKGWRLLAPITAYAOQTRGLLGCIIITSLTGRDKNOVEGEVQIVST 348  
 QY 54 AAQTFLATCINGCVTVYHGAGTRTIA SPKGPVIOMTYNDKDLVGPAPGQSRSLTPCT 113  
 DB 349 AAQTFLATCINGCVTVYHGAGTRTIA SPKGPVIOMTYNDKDLVGPAPGQSRSLTPCT 408  
 QY 114 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRISYLYKSGSGGLLCPAGHAVGIFRAAV 173  
 DB 409 CGSSDLXLVTRHADYIPVRRRGDSRGLSPRISYLYKSGSGGLLCPAGHAVGIFRAAV 468  
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 DB 469 CTRGVAKAVDFIPVESLETTMRSP 492

RESULT 11  
 AAP90158  
 ID AAP90158 standard; protein; 1786 AA.  
 XX  
 AC AAP90158;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 10-NOV-1989 (first entry)  
 XX  
 XX Protein sequence of hepatitis C virus composite cDNA.  
 DE  
 XX Hepatitis C virus; vaccine.  
 KW  
 OS Pan troglodytes.  
 XX  
 XX GB2212511-A.  
 XX  
 XX 26-JUL-1989.  
 PD  
 XX 18-NOV-1988; 88GB-0027024.  
 PF  
 XX 18-NOV-1987; 87US-0122714.  
 XX 30-DEC-1987; 87US-0139886.  
 PR  
 XX 26-FEB-1988; 88US-0161072.  
 PR  
 XX 26-OCT-1988; 88US-0263584.  
 PR

XX (CHIR ) CHIRON CORPORATION.  
 PA Houghton M, Choo QL, Kuo G;  
 XX  
 PI  
 XX  
 PT  
 PT  
 DR WPI: 1989-215054/30.  
 DR N-PSDB; AAN90327.  
 XX  
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment  
 PT of infection.  
 XX  
 XX Disclosure; fig 26; 30pp; English.  
 XX  
 CC The sequence is encoded by the composite cDNA of AAN90327. These  
 CC antigens react with antibodies in patients with non-A non-B hepatitis  
 CC (NANBH). They can be used to diagnose HCV-induced NANBH, to raise  
 CC antibodies for immunoassay or treatment, or to produce vaccines.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 XX Sequence 1786 AA;  
 SQ  
 Query Match 87.4%; Score 901.5; DB 10; Length 1786;  
 Best Local Similarity 85.8%; Pred. No. 7e-84;  
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;  
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCQETSOTGRDKNOVEGEVQIVST 53  
 Db 289 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLLCIIITSLTGRDKNOVEGEVQIVST 348  
 QY 54 AAQTFLATCINGVCWTVYHGAGTRTITASPKGPVIQMYTNVDKDLVGPAPQGSRLTPCT 113  
 Db 349 AAQTFLATCINGVCWTVYHGAGTRTITASPKGPVIQMYTNVDQDLVGPAPQGSRLTPCT 408  
 QY 114 CGSSDLVLTTRHADVTPVRRRGDSRGLSPRISYLVKSGSGPLLCPCAGHAGVIFRAAV 173  
 Db 409 CGSSDLVLTTRHADVTPVRRRGDSRGLSPRISYLVKSGSGPLLCPCAGHAGVIFRAAV 468  
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 Db 469 CTRGVAKAVDFIPVENLETTMRSP 492  
 RESULT 12  
 AAP90164  
 ID AAP90164 standard; protein; 2261 AA.  
 XX  
 AC AAP90164;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 01-NOV-1989 (first entry)  
 XX  
 XX Peptide encoded by composite hepatitis C virus cDNA.  
 XX  
 XX Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.  
 XX  
 XX Pan troglodytes.  
 XX  
 XX GB2212511-A.  
 XX  
 PD 26-JUL-1989.  
 XX  
 XX 18-NOV-1988; 88GB-0027024.  
 XX  
 XX 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORPORATION.  
 PA Houghton M, Choo QL, Kuo G;  
 XX  
 PI Houghton M, Choo QL, Kuo G;  
 XX  
 DR WPI: 1989-159274/22.  
 DR N-PSDB; AAN92103.  
 XX

DR WPI: 1989-215054/30.  
 DR N-PSDB; AAN90331.  
 XX  
 XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
 PT polypeptide(s) and antibodies for diagnosis, prevention and  
 PT treatment of infection.  
 XX  
 XX Disclosure; fig 32; 235pp; English.  
 XX  
 CC The sequence is the peptide encoded by the composite hepatitis C  
 CC virus (HCV) cDNA of AAN90331. The polypeptides are used to diagnose  
 CC HCV-induced NANBH, to raise antibodies for immunoassay or treatment,  
 CC or to produce vaccines.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 XX Sequence 2261 AA;  
 SQ  
 Query Match 87.4%; Score 901.5; DB 10; Length 2261;  
 Best Local Similarity 85.8%; Pred. No. 9.6e-84;  
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;  
 QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEGCQETSOTGRDKNOVEGEVQIVST 53  
 Db 380 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLLCIIITSLTGRDKNOVEGEVQIVST 439  
 QY 54 AAQTFLATCINGVCWTVYHGAGTRTITASPKGPVIQMYTNVDKDLVGPAPQGSRLTPCT 113  
 Db 440 AAQTFLATCINGVCWTVYHGAGTRTITASPKGPVIQMYTNVDQDLVGPAPQGSRLTPCT 499  
 QY 114 CGSSDLVLTTRHADVTPVRRRGDSRGLSPRISYLVKSGSGPLLCPCAGHAGVIFRAAV 173  
 Db 500 CGSSDLVLTTRHADVTPVRRRGDSRGLSPRISYLVKSGSGPLLCPCAGHAGVIFRAAV 559  
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 Db 560 CTRGVAKAVDFIPVENLETTMRSP 583  
 RESULT 13  
 AAP92047  
 ID AAP92047 standard; protein; 2301 AA.  
 XX  
 AC AAP92047;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 02-MAR-1990 (first entry)  
 XX  
 XX Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones  
 DE 12f through 15e.  
 XX  
 XX Hepatitis C virus (HCV); non-A, non-B hepatitis (NANBH).  
 XX  
 XX Hepatitis C virus.  
 XX  
 XX EP318216-A.  
 XX  
 XX 31-MAY-1989.  
 XX  
 XX 18-NOV-1988; 88EP-0310922.  
 XX  
 XX 18-NOV-1987; 87US-0122714.  
 PR 30-DEC-1987; 87US-0139886.  
 PR 26-FEB-1988; 88US-0161072.  
 PR 06-MAY-1988; 88US-0191263.  
 PR 26-OCT-1988; 88US-0263584.  
 PR 14-NOV-1988; 88US-0271450.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Houghton M, Choo QL, Kuo G;  
 XX  
 DR WPI: 1989-159274/22.  
 DR N-PSDB; AAN92103.  
 XX

XX Purified hepatitis C virus  
PT - and associated nucleic acids and polypeptide(s)  
XX  
XX Claim 13: Figure 32-1 - 32-7; 139 pp; English.  
PS

CC It is the sequence encoded in the open reading frame of hepatitis C virus  
CC (HCV) cDNA inserts in clones 12f through 15e. It is antigenic and could  
CC be used in immunoassay reagents and vaccines and to generate antibodies  
CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,  
CC non-B hepatitis.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)

[illegible]

RESULT 14	
AAP92050	
ID	AAP92050 standard; protein; 2436 AA.
XX	
AC	AAP92050;
XX	
AC	
XX	
DT	25-MAR-2003 (updated)
DT	02-MAR-1990 (first entry)
XX	
DE	Sequence encoded in the hepatitis C virus (HCV) cDNA inserts in clones
DE	K9-1 through 15e.
XX	
KW	Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH)
XX	
OS	Hepatitis C virus.
XX	
PN	EP318216-A.
XX	
PD	31-MAY-1989.
XX	
DF	18-NOV-1988; 88EP-0310922.
XX	
PR	18-NOV-1987; 87US-0122714.
PR	30-DEC-1987; 87US-0139886.
PR	26-FEB-1988; 88US-0161072.
PR	06-MAY-1988; 88US-0191263.
PR	26-OCT-1988; 88US-0263584.
PR	14-NOV-1988; 88US-0271450.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
PI	Houghton M, Choo QL, Kuo G;
XX	
DR	WPI; 1989-159274/22.
DR	N-PSDB; AAN92106.
XX	

PT	Purified hepatitis C virus
PT	- and associated nucleic acids and polypeptide(s)
XX	
PS	Claim 13; Figure 47-1 - 47-8; 139 pp: English.
XX	
CC	It is the sequence encoded in the open reading frame of hepatitis C virus
CC	(HCV) cDNA inserts in clones K9-1 through 15c. It is antigenic and could
CC	be used in immunosassay reagents and vaccines and to generate antibodies
CC	useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC	non-B hepatitis.
CC	(Updated on 25-MAR-2003 to correct PR field.)
CC	(Updated on 25-MAR-2003 to correct PI field.)
XX	
SQ	Sequence 2436 AA;

Query Match	87.4%	Score	901.5	DB	10	Length	2436
Best Local Similarity	85.8%	Pred. No.	1.le-83				
Matches	175	Conservative	9	Mismatches	11	Indels	9
Caps	1						
Qy	3	KKGSVVIVGRIN	-----LSGDTAYAQOTRCEGCQETSGTGRDKNKNOVEGVOIVST	53			
		::: ::::	::: ::::				
Db	555	RGREILLGPDAGMYSKGRLLA	ITAYAQOTRGLLCIITSITGRDKNKNOVEGVOIVST	614			
Qy	54	AAQTFLATCINGCVTVYHGAGTR	TIASPKGPVITOMYTNVDKDLGVHPAPQGSRSUTPCT	113			
Db	615	AAQTFLATCINGCVTVYHGAGTR	TIASPKGPVITOMYTNVDQDLGVHPAPQGSRSUTPCT	674			
Qy	114	CSSDDLVLVTRHADYIPVRRRGDS	RGSLSPRPISYLKSGSGGPLLCPAGHANGVIFRAAV	173			
Db	675	CSSDDLVLVTRHADYIPVRRGD	SRGSLSPRPISYLKSGSGGPLLCPAGHANGVIFRAAV	734			
Qy	174	CTRGAKAVDFIPVESLET	MRSP	197			
Db	735	CTRGAKAVDFIPVENLET	MRSP	758			

RESULT 15	
AAP90288	
ID	AAP90288 standard; protein; 2436 AA.
XX	XX
XX	AC
XX	AAP90288;
XX	XX
DT	25-MAR-2003 (updated)
DT	19-JUL-2001 (updated)
DT	01-NOV-1989 (first entry)
XX	XX
DE	Peptide encoded by composite hepatitis C cDNA.
XX	XX
KW	Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
XX	XX
OS	Pan troglodytes.
XX	XX
PN	GB2212511-A.
PD	XX
XX	26-JUL-1989.
XX	XX
PF	18-NOV-1988; 88GB-0027024.
XX	XX
PR	18-NOV-1987; 87US-0122714.
PR	30-DEC-1987; 87US-0139886.
PR	26-FEB-1988; 88US-0161072.
PR	26-OCT-1988; 88US-0263584.
XX	XX
PA	(CHIR ) CHIRON CORPORATION.
XX	XX
PI	Houghton M, Choo QL, Kuo G;
XX	XX
DR	WPI; 1989-215054/30.
DR	N-PSDB; AAP90336.
XX	XX
PT	Hepatitis C virus gene - used for prodn. of polynucleotide probes,
PT	polypeptide(s) and antibodies for diagnosis, prevention and
PT	treatment of infection.

```

XX PS Disclosure: fig 47-1 to 47-8; 235pp; English.
XX CC The sequence is the peptide encoded by the composite hepatitis C
CC virus (HCV) cDNA of AAN90336. The polypeptides are used to
CC diagnose HCV-induced NANBH, to raise antibodies for
CC immunosassay or treatment, or to produce vaccines.
CC (N.B. This record was resubmitted to correct errors in the sequence.)
CC (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 2436 AA;
SQ Query Match 87.4%; Score 901.5; DB 10; Length 2436;
Best Local Similarity 85.8%; Pred. NO. 1.1e-83;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;
QY 3 KKGSVVIVGRIN-----LSGDTAYAOQTRGEEGCOETSTQGRDKNQVEGEVOIVST 53
Db 555 RRGREILLGPADGMVSKGWRLLAPITAYAOQTRGLGCIITSLTGRDKNQVEGEVOIVST 614
QY 54 AAQOTFLATCINGVCVTVYHGAGTRTIA SPKGPVIQMYTNVDKDLVGPAPQGSRSLTPECT 113
Db 615 AAQOTFLATCINGVCVTVYHGAGTRTIA SPKGPVIQMYTNVDODLVGPAPQGSRSLTPECT 674
QY 114 CGSSDLXLVTRHADVIPVRRRGDSRGLLSRPISYLGSSGGPLLCPAGHAVGIFRAAV 173
Db 675 CGSSDLXLVTRHADVIPVRRRGDSRGLLSRPISYLGSSGGPLLCPAGHAVGIFRAAV 734
QY 174 CTRGVAKAVDFIPVESLETMTMRSP 197
Db 735 CTRGVAKAVDFIPVENLETMTMRSP 758
```

Search completed: August 30, 2003, 19:12:23  
Job time : 45.6227 secs





```

RESULT 3
GNWVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
C;Species: hepatitis C virus
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MOID:92052236; PMID:1658600
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <GPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:136,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240
Query Match 85.9% Score 886.5; DB 1; Length 3011;
Best Local Similarity 83.8%; Pred. No. 9, 2e-74;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;
Qy 3 KGSWTVGRIN-----LSGDTAYAOOTRGEGCOETSGTRDKNOVEGEVQIVST 53
:: : :: : | : ||||| : | : ||||| : ||||| : ||||| : ||||| :
Db 1005 RRGQEILLGPADGMVSKGWRLAPITAYAOOTRGLLCIITSLTGRDKNOVEGEVQIVST 1064
Qy 54 AAQTFLATCINGVCWTYYHGAGRTTASPKPGVPVQMVTYNDKDLVGWPAPQGSRSLTPCT 113
:: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1065 ATQTLATCINGVCWTYYHGAGRTTASPKPGVPVQMTYNDKDLVGWPAPQGSRSLTPCT 1124
Qy 114 CGSSDLYLVTRHADVIPRRRGDSRGLLSPRPISYLKSGSGGPLLCPCAGHAVGIFRAAV 173
:: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1125 CGSSDLYLVTRHADVIPRRRGDSRGLLSPRPISYLKSGSGGPLLCPTGHAVGLFAAAV 1184
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197
:: : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
RESULT 4
GNWVTH
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2261,2275,2301,2317,2323,2339,2345,2351,2367,2373,2389,2395,2401,2407,2413,2419,2425,2431,2437,2443,2449,2455,2461,2467,2473,2479,2485,2491,2497,2503,2509,2515,2521,2527,2533,2539,2545,2551,2557,2563,2569,2575,2581,2587,2593,2599,2605,2611,2617,2623,2629,2635,2641,2647,2653,2659,2665,2671,2677,2683,2689,2695,2701,2707,2713,2719,2725,2731,2737,2743,2749,2755,2761,2767,2773,2779,2785,2791,2797,2803,2809,2815,2821,2827,2833,2839,2845,2851,2857,2863,2869,2875,2881,2887,2893,2899,2905,2911,2917,2923,2929,2935,2941,2947,2953,2959,2965,2971,2977,2983,2989,2995,3001,3007,3013,3019,3025,3031,3037,3043,3049,3055,3061,3067,3073,3079,3085,3091,3097,3103,3109,3115,3121,3127,3133,3139,3145,3151,3157,3163,3169,3175,3181,3187,3193,3199,3205,3211,3217,3223,3229,3235,3241,3247,3253,3259,3265,3271,3277,3283,3289,3295,3301,3307,3313,3319,3325,3331,3337,3343,3349,3355,3361,3367,3373,3379,3385,3391,3397,3403,3409,3415,3421,3427,3433,3439,3445,3451,3457,3463,3469,3475,3481,3487,3493,3499,3505,3511,3517,3523,3529,3535,3541,3547,3553,3559,3565,3571,3577,3583,3589,3595,3601,3607,3613,3619,3625,3631,3637,3643,3649,3655,3661,3667,3673,3679,3685,3691,3697,3703,3709,3715,3721,3727,3733,3739,3745,3751,3757,3763,3769,3775,3781,3787,3793,3799,3805,3811,3817,3823,3829,3835,3841,3847,3853,3859,3865,3871,3877,3883,3889,3895,3901,3907,3913,3919,3925,3931,3937,3943,3949,3955,3961,3967,3973,3979,3985,3991,3997,4003,4009,4015,4021,4027,4033,4039,4045,4051,4057,4063,4069,4075,4081,4087,4093,4099,4105,4111,4117,4123,4129,4135,4141,4147,4153,4159,4165,4171,4177,4183,4189,4195,4201,4207,4213,4219,4225,4231,4237,4243,4249,4255,4261,4267,4273,4279,4285,4291,4297,4303,4309,4315,4321,4327,4333,4339,4345,4351,4357,4363,4369,4375,4381,4387,4393,4399,4405,4411,4417,4423,4429,4435,4441,4447,4453,4459,4465,4471,4477,4483,4489,4495,4501,4507,4513,4519,4525,4531,4537,4543,4549,4555,4561,4567,4573,4579,4585,4591,4597,4603,4609,4615,4621,4627,4633,4639,4645,4651,4657,4663,4669,4675,4681,4687,4693,4699,4705,4711,4717,4723,4729,4735,4741,4747,4753,4759,4765,4771,4777,4783,4789,4795,4801,4807,4813,4819,4825,4831,4837,4843,4849,4855,4861,4867,4873,4879,4885,4891,4897,4903,4909,4915,4921,4927,4933,4939,4945,4951,4957,4963,4969,4975,4981,4987,4993,4999,5005,5011,5017,5023,5029,5035,5041,5047,5053,5059,5065,5071,5077,5083,5089,5095,5101,5107,5113,5119,5125,5131,5137,5143,5149,5155,5161,5167,5173,5179,5185,5191,5197,5203,5209,5215,5221,5227,5233,5239,5245,5251,5257,5263,5269,5275,5281,5287,5293,5299,5305,5311,5317,5323,5329,5335,5341,5347,5353,5359,5365,5371,5377,5383,5389,5395,5401,5407,5413,5419,5425,5431,5437,5443,5449,5455,5461,5467,5473,5479,5485,5491,5497,5503,5509,5515,5521,5527,5533,5539,5545,5551,5557,5563,5569,5575,5581,5587,5593,5599,5605,5611,5617,5623,5629,5635,5641,5647,5653,5659,5665,5671,5677,5683,5689,5695,5701,5707,5713,5719,5725,5731,5737,5743,5749,5755,5761,5767,5773,5779,5785,5791,5797,5803,5809,5815,5821,5827,5833,5839,5845,5851,5857,5863,5869,5875,5881,5887,5893,5899,5905,5911,5917,5923,5929,5935,5941,5947,5953,5959,5965,5971,5977,5983,5989,5995,6001,6007,6013,6019,6025,6031,6037,6043,6049,6055,6061,6067,6073,6079,6085,6091,6097,6103,6109,6115,6121,6127,6133,6139,6145,6151,6157,6163,6169,6175,6181,6187,6193,6199,6205,6211,6217,6223,6229,6235,6241,6247,6253,6259,6265,6271,6277,6283,6289,6295,6301,6307,6313,6319,6325,6331,6337,6343,6349,6355,6361,6367,6373,6379,6385,6391,6397,6403,6409,6415,6421,6427,6433,6439,6445,6451,6457,6463,6469,6475,6481,6487,6493,6499,6505,6511,6517,6523,6529,6535,6541,6547,6553,6559,6565,6571,6577,6583,6589,6595,6601,6607,6613,6619,6625,6631,6637,6643,6649,6655,6661,6667,6673,6679,6685,6691,6697,6703,6709,6715,6721,6727,6733,6739,6745,6751,6757,6763,6769,6775,6781,6787,6793,6799,6805,6811,6817,6823,6829,6835,6841,6847,6853,6859,6865,6871,6877,6883,6889,6895,6901,6907,6913,6919,6925,6931,6937,6943,6949,6955,6961,6967,6973,6979,6985,6991,6997,7003,7009,7015,7021,7027,7033,7039,7045,7051,7057,7063,7069,7075,7081,7087,7093,7099,71

Db 1065 ATQSFATCVNGVWTVYHAGSKTLAGPGPIITQMYTNVDQDLVGPAPGARSMPCT 1124  
Qy 114 CGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184  
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 7  
A:Variety: isolate JK1  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: S18030  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa  
A:Reference number: S18028  
A:Accession: S18030  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <HON>  
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat  
A:Reference number: A48332; MUID:93119270; PMID:8380322  
A:Accession: S33570  
A:Molecule type: genomic RNA  
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
as Trp, and TTC for residue 771 as Ser  
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1317/Product: hepatitis C virus genome polyprotein  
F:1230-1317/Product: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 81.8%; Score 844.5; DB 1; Length 3010;  
Best Local Similarity 77.5%; Pred. No. 7.4e-70;  
Matches 158; Conservative 20; Mismatches 17; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAAQTGREGGCGTSTQTRDKNOVEGEVQIVST 53  
Db 1005 RRGREILLGPADSIQGGWRLAPITAYAAQTGRLGCVITSLTGRDKNOVEGEVQIVST 1064  
Qy 54 AAQTFLATCINGVWTVYHAGSKTLAGPGPIITQMYTNVDQDLVGPAPGARSMPCT 113  
Db 1065 ATQSFATCVNGVWTVYHAGSKTLAGPGPIITQMYTNVDQDLVGPAPGARSMPCT 1124  
Qy 114 CGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184  
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 8  
S18030  
genome polyprotein - hepatitis C virus (isolate JK1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: S18030  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina  
A:Reference number: JC5620; MUID:97366593; PMID:9223423

A:Variety: isolate JK1  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C:Accession: S18030; S33570; A48332; S18029  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa  
A:Reference number: S18028  
A:Accession: S18030  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <HON>  
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat  
A:Reference number: A48332; MUID:93119270; PMID:8380322  
A:Accession: S33570  
A:Molecule type: genomic RNA  
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
as Trp, and TTC for residue 771 as Ser  
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:116-191/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1230-1317/Product: hepatitis C virus genome polyprotein  
F:1230-1317/Product: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 80.3%; Score 828.5; DB 1; Length 3010;  
Best Local Similarity 76.5%; Pred. No. 2.3e-68;  
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAAQTGREGGCGTSTQTRDKNOVEGEVQIVST 53  
Db 1005 RRGREILLGPADSIQGGWRLAPITAYAAQTGRLGCVITSLTGRDKNOVEGEVQIVST 1064  
Qy 54 AAQTFLATCINGVWTVYHAGSKTLAGPGPIITQMYTNVDQDLVGPAPGARSMPCT 113  
Db 1065 ATQSFATCVNGVWTVYHAGSKTLAGPGPIITQMYTNVDQDLVGPAPGARSMPCT 1124  
Qy 114 CGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 173  
Db 1125 YGSSDLVLTTRHADVPVRRGRDGRGSLSPRPISYLGKSGGGLPCPAGHAVGIFRAAV 1184  
Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTRSP 1208

RESULT 9  
JC5620  
genome polyprotein - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: JC5620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina  
A:Reference number: JC5620; MUID:97366593; PMID:9223423

A:Accession: JCS5620  
A:Molecule type: mRNA  
A:Residues: 1-3014 <CHA>  
A:Cross-references: GB:Y13184  
A:Experimental source: Isolate HC-J6, which predominates in South Africa  
A:Note: The translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:384-408/Region: hypervariable #status predicted  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEXH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 74.3%; Score 766.5; DB 1; Length 3014;  
Best Local Similarity 69.6%; Pred. No. 1.3e-62;  
Matches 142; Conservative 25; Mismatches 28; Indels 9; Gaps 1;

QY 3 KKGSVVIGRIN-----LSGDTAYAQOTRGEBCQETSGTGRKKNQVEGVQIVST 53  
DB 1006 RGRREIFLGPADIKTSGWRLAPITAYAQOTRGVUGAIVLSLTGRKNEAGEVQFLST 1065  
QY 54 AAQTFLATCINGCVTVYHGAGTRTIIASPKGVQIOMYTNVDKDLGWPAPOGSRSLTPCT 113  
DB 1066 ATOTFLGICINGVMVTLFGAGSKTLAGPKPVQVQMYTNVDKDLGWMPSPPKGSLTRCT 1125  
QY 114 CGSSDLYLTRADYVTPVRRGDSRGLSPRIISYLVKSSGGPCLPCAGHAGVIFRAAV 173  
DB 1126 CGSADLYLTRADYVTPVRRGDSRGLSPRIISYLVKSSGGPCLPCAGHAGVIFRAAV 1185  
QY 174 CTRGVAKAVDFIPVESLETMRSP 197  
DB 1186 CTRGVAKALEFVVENLETMRSP 1209

RESULT 10  
JQ1303  
genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
N:Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
A:Reference number: JQ1303  
A:Accession: JQ1303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651  
A:Experimental source: Isolate HC-J6 from a Japanese individual  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis C virus genome polyprotein NS3  
F:1316-1321/Region: nucleotide-binding motif A (P-loop)  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 65.1%; Score 672; DB 1; Length 3033;  
Best Local Similarity 68.7%; Pred. No. 8.4e-54;  
Matches 123; Conservative 24; Mismatches 32; Indels 0; Gaps 0;

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2031

Query Match 65.3%; Score 674; DB 1; Length 3033;  
Best Local Similarity 68.2%; Pred. No. 5.5e-54;  
Matches 122; Conservative 26; Mismatches 31; Indels 0; Gaps 0;

QY 19 TAYAQOTRGEBCQETSGTGRKKNQVEGVQIVSTAAQTFLATCINGCVTVYHGAGTRT 78  
DB 1034 TAYAQOTRGLLGTIVVMTGRDKTQAGEIQVLSIVTOSFLGTTISGVLTVYHGAGNKT 1093  
QY 79 IASPKGPVQIOMYTNVDKDLGWPAPOGSRSLTPCTCGSSDLYLTRADYVTPVRRGDSR 138  
DB 1094 LAGSRGPVTOMYSSAEGDLGVWSPSPGPKSLPCTCGAVDLYLVTRNADVPARRGDKR 1153  
QY 139 GSLLSPRLSYLVKSSGGPCLPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRSP 197  
DB 1154 GALLSPRLSTLKGSSGGPVLCPRGHAGVIFRAAVCSRGVAKSIDFIPVETLDIVTRSP 1212

RESULT 11  
GNVJ78  
genome polyprotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
N:Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
A:Accession: A40250; PQ0357; PQ0559  
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, Y.  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to  
A:Reference number: A40250; MUID:92230232; PMID:1314459  
A:Accession: A40250  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0397  
A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D10134  
A:Experimental source: isolate E-b12  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotoh  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: PQ0554; MUID:92068204; PMID:1720309  
A:Accession: PQ0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis C virus genome polyprotein NS3  
F:1316-1321/Region: nucleotide-binding motif A (P-loop)  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,203



A:Cross-references: GDB:4573993  
A:Map position: l4q32.1-l4q32.1  
C:Superfamily: unassigned WD repeat proteins; WD repeat homology  
F:148-181/Domain: WD repeat homology <WD1>  
F:414-447/Domain: WD repeat homology <WD2>

Query Match 7.9%; Score 82; DB 2; Length 452;  
Best Local Similarity 25.3%; Pred. No. 7.1;  
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;

QY	68	WTVVHGACTRTIASPKGPIOMYTINVDKDLVGPAPQGSRSLSL-----TPCTCGSSDLYLV	122
Db	197	WAEWH-----PRAPVLAGT-ADGNTWVKVPNGDKTFOGPNCPATCGR-----	240
QY	123	TRHADVIPVRRR---GDSRGS-----LLSPRPISYLGSSG--GPLLCPA-----	162
Db	241	-----VLPDGKRAVVGIEDGTIRIWLKQGSPIHVLKGTGEGHGLTCVAAANQDGLILT	295
QY	163	-----GHAVGIFR-----AAVCTRGVAKAVDFIPVESL	190
Db	296	GSVDCQAKLVSATCKVVGFRPETVASQPSLGECESESNSVESL	341

Search completed: August 30, 2003, 19:20:27  
Job time : 17.2134 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	901.5	87.4	3011	1 POLG_HCV1	P26664 h genome po
2	886.5	85.9	3011	1 POLG_HCVH	P27958 h genome po
3	854.5	82.8	3010	1 POLG_HCVTW	P29846 h genome po
4	850.5	82.4	3010	1 POLG_HCVBK	P26663 h genome po
5	850.5	82.4	3010	1 POLG_HCVJA	P26662 h genome po
6	844.5	81.8	3010	1 POLG_HCVJT	Q00269 h genome po
7	674	65.3	3033	1 POLG_HCVJ6	P26660 h genome po
8	672	65.1	3033	1 POLG_HCVJ8	P26661 h genome po
9	85.5	8.3	485	1 Y136_TREPA	O83172 treponema p
10	85	8.2	321	1 HHOA_ARATH	Q9sel7 arabidopsis
11	82	7.9	452	1 AAMP_HUMAN	Q13685 homo sapien
12	80.5	7.8	437	1 DEGL_ARATH	O22609 arabidopsis
13	78.5	7.6	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
14	77.5	7.5	2663	1 CENE_HUMAN	Q02224 homo sapien
15	76	7.4	911	1 TBIL_NEIMS	Q09056 neisseria m
16	75.5	7.3	786	1 SNIL_HUMAN	P57059 homo sapien
17	75	7.3	603	1 ENV_RSVP	P33396 rous sarcom
18	75	7.3	1705	1 TPPO_MOUSE	P70289 mus musculu
19	74.5	7.2	415	1 ZP3_RABIT	P48833 oryctolagu
20	74.5	7.2	776	1 HYPF_AZOVI	P40596 azotobacter
21	74	7.2	1165	1 POL_GALV	P21414 gibbon ape
22	73.5	7.1	263	1 GRAB_MOUSE	O35205 mus musculu
23	73.5	7.1	661	1 INV8_DAUCA	P80065 daucus caro
24	73	7.1	253	1 CAC3_BOVIN	P05805 bos taurus
25	73	7.1	259	1 IBPL_HUMAN	P08833 homo sapien
26	72.5	7.0	257	1 GRAM_HUMAN	P51124 homo sapien
27	72.5	7.0	706	1 TRFE_HORSE	P74225 equus cabal
28	72	7.0	659	1 VST2_HEVNE	Q03500 hepatitis e
29	72	7.0	1527	1 CALH_MOUSE	P39061 mus musculu
30	71.5	6.9	248	1 GRAD_MOUSE	P11033 mus musculu
31	71.5	6.9	248	1 TRY1_CHICK	Q90627 gallus gall
32	71.5	6.9	408	1 SEPR_THESR	P80146 thermus sp.
33	71	6.9	336	1 UL16_EBV	P03221 epstein-bar

34	71	6.9	844	1 CN4A_RAT	P54748 rattus norv
35	71	6.9	915	1 TBPI_NEIGO	Q01996 neisseria g
36	71	6.9	1180	1 ITAL_RAT	P18614 rattus norv
37	71	6.9	1240	1 YQU3_CAEEL	Q09550 caenorhabdl
38	70.5	6.8	264	1 CTRL_HUMAN	P40313 homo sapien
39	70.5	6.8	443	1 FLIT_AQUAE	P08531 aquifex aeo
40	70.5	6.8	642	1 ENV_FLVGL	P08359 feline leuk
41	70.5	6.8	660	1 VST2_HEVBU	P29326 hepatitis e
42	70.5	6.8	660	1 VST2_HEVPA	P33426 hepatitis e
43	70.5	6.8	743	1 TFE3_HUMAN	P19532 homo sapien
44	70	6.8	326	1 PANE_RHILO	Q987H5 rhizobium l
45	70	6.8	349	1 TRPD_PSEPU	P20575 pseudomonas

#### ALIGNMENTS

RESULT 1  
POLG\_HCV1  
ID POLG\_HCV1 STANDARD; PRT; 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DI 01-AUG-1992 (Rel. 23, Last sequence update)  
DI 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);  
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate 1) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,  
Bradley D.W., Kuo G., Houghton M.;  
RA \*Genetic organization and diversity of the hepatitis C virus.\*;  
Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
RL -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
precursor polyprotein, commonly with Asp or Glu in the P6  
position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
(RNA)(N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND RNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
-----

EMBL; M62321; AAA45676.1; -  
PIR; A39166; GNVVC3.  
PDB; 1ALV; 16-FEB-99.  
PDB; 1HEI; 25-NOV-98.  
MEROPS; S29.001; -  
MEROPS; S39.001; -  
InterPro; IPR001410; DEAD.  
InterPro; IPR002522; HCV\_capsid.



DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXdc; 1.  
 DR PolyProtein: Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 192 383 MATRIX ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS3 (POTENTIAL).  
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT SITE 1316 1319 ATP (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 65P8C9447FCESAF9 CRC64;

Query Match 37.48; Score 901.5; DB 1; Length 3011;  
 Best Local Similarity 85.84; Pred. No. 5.5e-77;  
 Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;  
 3 KKGSVIVGRIN-----LSGDTAYAAQOTRCEGCQETSQTGRDKNOVEGEQIVST 53

DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLCITSLTGRDKNOVEGEQIVST 1064  
 QY 54 AAQTFLATCINGVCWTVYHGAGTETIASPKGPVQIYNTVDKDLVGWPAQGSRLTPTCT 113  
 DB 1065 AAQTFLATCINGVCWTVYHGAGTETIASPKGPVQIYNTVDKDLVGWPAQGSRLTPTCT 1124  
 QY 114 CGSSDLVLTBHADVIVRRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAVGIFRAAV 173  
 DB 1125 CGSSDLVLTBHADVIVRRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAVGIFRAAV 1184  
 QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=111108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 virus: comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 oligonucleotide: the crystal structure provides insights into the mode  
 of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.



CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M67463; AAA45534.1; -  
 CC PIR: A36814; GNMVCH.  
 CC PDB: 1HEI; 25-NOV-98.  
 CC PDB: 1AIV; 16-FEB-99.  
 CC PDB: 1A1R; 17-JUN-98.  
 CC MEROPS: S29.001; -  
 CC MEROPS: U39.001; -  
 CC TRANSFAC: T04155; -  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV Capsid.  
 CC InterPro: IPR002521; HCV core.  
 CC InterPro: IPR002519; HCV env.  
 CC InterPro: IPR002531; HCV NS1.  
 CC InterPro: IPR002518; HCV NS2.  
 CC InterPro: IPR004109; HCV NS3.  
 CC InterPro: IPR000745; HCV NS4a.  
 CC InterPro: IPR001490; HCV NS4b.  
 CC InterPro: IPR002868; HCV NS5a.  
 CC InterPro: IPR002166; HCV RdRp.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC InterPro: IPR007095; RNA\_pol\_DS\_Ps.  
 CC InterPro: IPR007094; RNA\_pol\_PSVir.  
 CC Pfam: PF01543; HCV\_capsid; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01539; HCV\_env; 1.  
 CC Pfam: PF01560; HCV\_NS1; 1.  
 CC Pfam: PF01538; HCV\_NS2; 1.  
 CC Pfam: PF02907; HCV\_NS3; 1.  
 CC Pfam: PF01006; HCV\_NS4a; 1.  
 CC Pfam: PF01001; HCV\_NS4b; 1.  
 CC Pfam: PF01506; HCV\_NS5a; 1.  
 CC Pfam: PF00271; Helicase\_C; 1.  
 CC Pfam: PF00998; Viral\_RdRp; 1.  
 CC ProDom: PD186062; HCV\_NS1; 1.  
 CC SMART: SM00487; DEXDC; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 KW INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 1 191  
 FT CHAIN 192 383 ENVELOPE GLYCOPROTEIN E1.  
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E2.  
 FT CHAIN 747 809 PROTEIN P7.  
 FT CHAIN 810 1026 NONSTRUCTURAL PROTEIN NS2.  
 FT CHAIN 1027 1657 PROTEASE/HELICASE NS3.  
 FT CHAIN 1658 1711 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.  
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.  
 FT CHAIN 347 369 POTENTIAL.  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH\_BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT 305 305 CARBOHYD  
 FT 417 417 CARBOHYD  
 FT 423 423 CARBOHYD  
 FT 430 430 CARBOHYD  
 FT 448 448 CARBOHYD  
 FT 476 476 CARBOHYD  
 FT 532 532 CARBOHYD  
 FT 540 540 CARBOHYD  
 FT 556 556 CARBOHYD  
 FT 576 576 CARBOHYD  
 FT 623 623 CARBOHYD  
 FT 645 645 CARBOHYD  
 FT 1224 1224 STRAND  
 FT 1232 1232 TURN  
 FT 1238 1238 TURN  
 FT 1239 1239 TURN  
 FT 1246 1246 HELIX  
 FT 1247 1247 TURN  
 FT 1251 1251 STRAND  
 FT 1258 1258 HELIX  
 FT 1271 1271 TURN  
 FT 1272 1272 TURN  
 FT 1277 1277 STRAND  
 FT 1281 1281 TURN  
 FT 1282 1282 TURN  
 FT 1283 1283 STRAND  
 FT 1291 1291 STRAND  
 FT 1296 1296 HELIX  
 FT 1302 1302 TURN  
 FT 1312 1312 STRAND  
 FT 1317 1317 TURN  
 FT 1323 1323 HELIX  
 FT 1335 1335 TURN  
 FT 1336 1336 TURN  
 FT 1343 1343 STRAND  
 FT 1352 1352 TURN  
 FT 1360 1360 TURN  
 FT 1362 1362 STRAND  
 FT 1368 1368 STRAND  
 FT 1373 1373 STRAND  
 FT 1375 1375 TURN  
 FT 1376 1376 TURN  
 FT 1378 1378 STRAND  
 FT 1382 1382 HELIX  
 FT 1389 1389 STRAND  
 FT 1393 1393 STRAND  
 FT 1397 1397 HELIX  
 FT 1410 1411 TURN  
 FT 1414 1414 STRAND  
 FT 1419 1420 TURN  
 FT 1432 1436 STRAND  
 FT 1438 1439 TURN  
 FT 1450 1453 STRAND  
 FT 1456 1463 STRAND  
 FT 1471 1478 STRAND  
 FT 1480 1480 STRAND  
 FT 1481 1488 HELIX  
 FT 1489 1490 TURN  
 FT 1497 1501 STRAND  
 FT 1507 1507 STRAND  
 FT 1511 1511 STRAND  
 FT 1514 1527 HELIX  
 FT 1532 1544 HELIX  
 FT 1550 1550 STRAND  
 FT 1555 1564 HELIX  
 FT 1570 1578 HELIX  
 FT 1579 1580 TURN  
 FT 1584 1597 HELIX  
 FT 1598 1598 TURN  
 FT 1606 1611 HELIX  
 FT 1614 1618 TURN  
 FT 1622 1623 STRAND  
 FT 1627 1627 STRAND  
 FT 1635 1636 STRAND  
 FT 1640 1652 HELIX  
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;  
 Query Match 85.9%; Score 886.5; DB 1; Length 3011;  
 Best Local Similarity 83.8%; Pred. No. 1.5e-75;

```
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;
QY 3 KGSVIVIGRIN-----LSGDTAYAOOTREGCQTSOTGRKKNQVEGEVIVST 53
Db 1: : : : :
1005 RRGQILLGPADGMVSKGWELLAPITAYAOOTREGLLCIIITSLTGRKKNQVEGEVIVST 1064
QY 54 AAOITFLATCINGVWTVYHCGAGTRTITASPKGVPVQMYTNVDKLVGMPAPQGSRSITPCT 113
Db 1: : : : :
1065 ATOITFLATCINGVWTVYHCGAGTRTITASPKGVPVQMYTNVDKLVGMPAPQGSRSITPCT 1124
QY 114 CGSDDLVLVTRHADVIPVRRGDSRGLSPRISYLGSSGGLLCPAGHAYGIFRAAV 173
Db 1: : : : :
1125 CGSDDLVLVTRHADVIPVRRGDSRGLSPRISYLGSSGGLLCPCTCHAVGLFRAAV 1184
QY 174 CTRGVAKAVDIPVESLETHRSP 197
Db 1: : : : :
1185 CTRGVAKAVDIPVENLETHRSP 1208

RESULT 3
POLG_HCVTW
ID POLG_HCVTW STANDARD; PRT: 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; PubMed=1314449;
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}[N].
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84754; ; NOT_ANNOTATED_CDS.
DR PIR: A40244; GNAVTV.
DR PDB: 1M64; 25-FEB-03.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
```

```
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CORE PROTEIN (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
```

Query Match 82.8%; Score 854.5; DB 1; Length 3010;  
Best Local Similarity 78.9%; Pred. No. 1.6e-72;  
Matches 161; Conservative 18; Mismatches 16; Indels 9; Gaps 1;

```

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOETOTGRKKNQVGEVQIVST 53
DB 1005 RRGREILLGPADECKGKRLNLLAPITAYAAQOTRGLFGCIITSLTGRKKNQVGEVQIVST 1064
QY 54 AAOFTFLATCINGVCTVYVYGAGTRITITASPKGPVQIOWTWDKDLVQWPAQGSRLTPTCT 113
DB 1065 ATOSFLATCINGVCTVYVYGAGSKTLGAGPKGPTITOMYTNDQDLVGVHAPQAGARSILTPTCT 1124
QY 114 CGSSDLYLVTRHADVTPVRRRGRSGSLLSPRISYLKSGSGPGLLCPAGHAGVGFRAAV 173
DB 1125 CGSSDLYLVTRHADVTPVRRRGRSGSLLSPRISYLKSGSGPGLLCPCHVGVGFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 4
POLG_HCVBK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RN MEDLINE-91140696; PubMed=1847104;
RX Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RX Onishi E., Andoh T., Yoshida I., Okayama H.;
RA "Structure and organization of the hepatitis C virus genome isolated
RA from human carriers.";
RT J. Virol. 65:1105-1113(1991).
RN [2]
RN SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moosaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +

```

```

[RNA](N).
-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58335; AAA72945.1;
DR PIR; A38465; GNWVTC.
DR PDB; 1AIQ; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR PDB; 1NS3; 08-APR-98.
DR PDB; 1C2P; 15-NOV-00.
DR PDB; 1CSJ; 08-NOV-99.
DR PDB; 1GX5; 09-APR-02.
DR PDB; 1GA6; 10-APR-02.
DR PDB; 1QTV; 26-JUN-00.
DR PDB; 8OHM; 20-APR-99.
DR MEROPS; S29.001;
DR MEROPS; U39.001;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 131 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).

```

```

FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1031 1035 N-LINKED (GLCNAC. .) (POTENTIAL).
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT STRAND 1101 1103
FT TURN 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1680 1688
SQ SEQUENCE 3010 AA: 327189 MW: P8422D5ECCDFD9C CRC64;

Query Match 82.4%; Score 850.5; DB 1; Length 3010;
Best Local Similarity 77.5%; Pred. No. 3.8e-72;
Matches 158; Conservative 21; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQTRGEGCQETISQTRKDKQVEGEVQIVST 53
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 1005 RRGKEILLGPADSLGRLGLLAPITAYSQOTRGLGCIITSLTGRDKQVEGEVQIVST 1064
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 54 AATFATLCTGVCVTVYHAGAGTRTASPKGPVQIYTWVDKDLVGVMPAPQSGRSITPCT 113
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 1065 ATQSFLATCTGVCVTVYHAGAGSKTLAAPKGPITQYTWVDQDLVGVMPKPGARSITPCT 1124
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 114 CGSSDLYLTHRHADVTPVRRGDSRGLSPRISYLKSGSGGPLLCAGHAGVIFRAAV 173
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 1125 CGSSDLYLTHRHADVTPVRRGDSRGLSPRISYLKSGSGGPLLCAGHAGVIFRAAV 1184
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
   | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

```

Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90208; BAA14233.1; -.
DR PIR: A39253; GNWVCJ.
DR HSSP: P26663; LXP.
DR MEROPS: S29.001; -.
DR MEROPS: S29.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.

```

DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Coat protein; Hydrolyase; Serine protease.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327017 MW; AA593794F46DB185 CRC64;

Query Match 82.4%; Score 850.5; DB 1; Length 3010;  
 Best Local Similarity 76.5%; Pred. No. 3.8e-72;  
 Matches 156; Conservative 23; Mismatches 16; Indels 9; Gaps 1;

Qy 3 KGSWVIVGRINLSGD-----TAYAGTRGEGCGEOSTGTGRDNKNOVEGVQIVST 53  
 Db 1005 RGKEILLGADSGEGCGWRLLAPITAYSOQTRGLLCITITSTGRDNKNOVDGEVQLST 1064

Qy 54 AAQFLATFCINGVCWTVHGGAGTGTATSPKGPVITQMTYNDKDLVGPAPQSGRSITPCT 113  
 Db 1065 ATQSFATFCVNGVCWTVHGGAGTGTATSPKGPVITQMTYNDKDLVGPAPQSGRSITPCT 1124

Qy 114 CGSSDLYLIVTHADYIVPVRGRDGRGSLSPRISYILKSGSGGGLLCPAGHANGVIFRAAV 173  
 Db 1125 CGSSDLYLIVTHADYIVPVRGRDGRGSLSPRISYILKSGSGGGLLCPAGHANGVIFRAAV 1184

Qy 174 CTRGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 6

ID POLG\_HCVJT STANDARD; PRT; 3010 AA.  
 AC Q00269;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-JT) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=31642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92295714; PubMed-1318627;  
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
 RT carrier: sequence variation within the same individual and among  
 RT infected individuals";  
 RL Virus Res. 23:39-53(1992).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D11168; BAA01943.1; .  
 DR PIR: A45573; A45573  
 DR PDB: 1AIQ; 25-MAR-98.  
 DR PDB: 1JXP; 14-JAN-98.  
 DR MEROPS: S29.001; .  
 DR MEROPS: U39.001; .  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.







DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;  
 Query Match 65.1%; Score 672; DB 1; Length 3033;  
 Best Local Similarity 68.7%; Pred. No. 3.1e-55;  
 Matches 123; Conservative 24; Mismatches 32; Indels 0; Gaps 0;  
 QY 19 TAYAQOTRGECCQETSGTGRDKNQVEGEVOIVSTAQTFLATCINGVCHVTHGAGTRT 78  
 DB 1034 TAYTQOTRGLLGAIVVSLTGRDKNEAQGVQLSSVTQTFLGTSISGLVLTVTHGAGNKT 1093  
 QY 79 IASPKGPVIOYTNVDKLVGPAQGSRSUTPCTCGSSDLVLTTRADVTPVRRGDSR 138  
 DB 1094 LAGPKGPVIOYTNVDKLVGPAQGSRSUTPCTCGSSDLVLTTRADVTPVRRGDSR 1153  
 QY 139 GSLSPRISYLKSSCGPLCPAGHVGIPRAAVCTRGNAKAVDFIPVSELTMRSP 197  
 DB 1154 GALLSPRLSLTKGSGGVLCSRGHVGIPRAAVCTRGNAKAVDFIPVSELTMRSP 1212  
 RESULT 9  
 Y136\_TREPA  
 ID Y136\_TREPA STANDARD; PRT; 485 AA.  
 AC O83172;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical lipoprotein TP0136 precursor.  
 GN TP0136.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxId=160;

[1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khatak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete";  
 RL Science 281:375-388(1998).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 CC (potential).  
 CC -!- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: AB001199; AAC65137.1; ALT\_INIT.  
 DR TIGR: TP0136; -  
 KW Hypothetical protein; Lipoprotein; Membrane; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.  
 FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).  
 FT DOMAIN 164 178 GLY/SER-RICH.  
 FT DOMAIN 196 210 GLY/SER-RICH.  
 FT DOMAIN 253 267 GLY/SER-RICH.  
 FT DOMAIN 318 327 POLY-SER.  
 FT DOMAIN 444 447 POLY-SER.  
 SQ SEQUENCE 485 AA; 48984 MW; C7A4CEDC7DC5CED CRC64;  
 Query Match 8.3%; Score 85.5; DB 1; Length 485;  
 Best Local Similarity 23.4%; Pred. No. 1.3;  
 Matches 50; Conservative 17; Mismatches 76; Indels 71; Gaps 10;  
 QY 16 SGTAYV-----OQTRGEGCQETSGTGRDKNQVEGEVOIVSTAQTFLATCI- 63  
 DB 54 AGSKLYATNCRWLWELNLTGSGWQVSSSVPTDSK-----KVSATDGTFTVLACVP 108  
 QY 64 -NGCVTVTHGAG---TRTIAPKGPVIOYTNVDKLVG-----WPAQGSRSLTPTCT 113  
 DB 109 GTGVYKHCYVNGAGSSSTGTGTASPTSTCQHAT-----LVGGTSKPFVLVPGGTGNGMCG 164  
 QY 114 C-----GSSDLYLVTHADVIP-----VRRRGDSRGLSPRPISYLK----- 151  
 DB 165 CGGGGGGSSSSSSCHIIWLVPFGTNGNCGCGGGGGSSSSSSCHIKVNTDEOFL 224  
 QY 152 -----GSSGGLPLCPAGHVG 167  
 DB 225 DMGEYVVVTKHLYTKNGSSSAGAACQCGGGGG 258  
 RESULT 10  
 HHOA\_ARATH  
 ID HHOA\_ARATH STANDARD; PRT; 321 AA.  
 AC O9SEL7; O49507;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).  
 GN HHOA OR AT4G18370 OR F28J12.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;







```

Db 150 VPOGSGFWMQKQGHVTHYHVRGASDLRVTLAQDTTDAKVVGFDQDKDVAVLRLDA 209
QY 103 PGGSRLTPTCTGSSDLXLY-----TRHADVTPVRRRGDSRGSLLSPRP 147
Db 210 PK--NKLRPVGVSAADLVGVKQVFAIGNPFLDHTLTIGVSGLRREIS--SAATGRPI 265
QY 148 SYL-----KGSGGGLPCLCPAGHAGVGFRAAVCTRGVAKAVDF-IPVESL 190
Db 266 QDVQIOTDAAINFGNSGGPLDSSGTLIGINTAIYSPGASSGVGFSIPVDV 317

RESULT 13
PAAD_PSEAE STANDARD; PRT; 209 AA.
AC Q9HX08; 201 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aromatic acid decarboxylase (EC 4.1.1.-).
GN PA4019.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy S.N., Tolentino E.R., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE /
CC PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004818; AAC07406.1;
CC PIR; H83144; H83144.
CC InterPro; IPR003382; Flavoprotein.
CC Pfam; PF02441; Flavoprotein; 1.
CC KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.
CC SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;

Query Match 7.68; Score 78.5; DB 1; Length 209;
Best Local Similarity 26.5%; Pred. No. 2.3;
Matches 50; Conservative 16; Mismatches 56; Indels 67; Gaps 11;

QY 43 QVEGEVO-IVSTAOTFLATCINGVCVTYVHGACTRTTASPKGP----- 85
Db 29 QEEREVFLISKAAQLVWAT-----ETDVALPAKPAQMAQFLTEYCGAAG 74
QY 86 VIQMTYNDKDLVGNPAGQSRSLTP-----CTGSSDL-----YLVTHADYVPI 131
Db 75 QIRVFQND-----WAPPASGASPNAMVICPSTGTLTSAVATGACNNLIERAADVALK 129
QY 132 RRGDSRGSLLSPR--PIS-----YKSGSGGPLLCPCAGHAGVGFRAAVCTRGVAKAVD 193
Db 130 ER----RPLVVPREPFPSSITHLENLKLNLGAVILPA--AFGFYHQ----POSVEDLVD 180
QY 184 FIPVESLET 192

```

```

Db 181 FVVARILNT 189

RESULT 14
CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RL microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z15005; CAA78727.1;
CC PIR; S28261; S28261.
CC HSP; P17119; 3KAR.
CC Genew; HGNC:1856; CENPE.
CC GK; Q02224;
CC MIM; 117143;
CC GO; GO:0005699; C:kinetochore; TAS.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0008350; F:kinetochore motor activity; TAS.
CC GO; GO:0000067; F:DNA replication and chromosome cycle; TAS.
CC GO; GO:0007079; P:mitotic chromosome movement; TAS.
CC GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

```

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
KW Cell cycle; Centromere.  
FT DOMAIN 1 335 KINESIN-MOTOR.  
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
FT NP\_BIND 86 93 ATP (BY SIMILARITY).  
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 7.5%; Score 77.5; DB 1; Length 2663;  
Best Local Similarity 24.0%; Pred. No. 56;  
Matches 41; Conservative 15; Mismatches 52; Indels 63; Gaps 8;  
QY 32 QETSGTRDKNQVEGEQIVSTAAQT-----LATCINGVCWTVYHGAGRTIA 80  
Db 2523 CQNEQLIKQKNELLNNOHLSNEVTKWTKRLKRAHKQVTCE----- 2566  
QY 81 SPKGPVIOYNTVDKDLVCHVPAQGSRLTPCTCGSSDLVLTTRHADVIPVRRG-----D 136  
Db 2567 SPKSPKVTGTASKKK-----OITPSOCKERNL-----QDPVKPSKSCFFD 2608  
QY 137 SRG-SLLSPRISYLKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIP 186  
Db 2609 SRSKSLSPSPHVRFDNSSLG--LCPEVQNAG-----AESVDSQP 2646

RESULT 15  
TB1L\_NEIMB  
ID TB1L\_NEIMB STANDARD; PRT: 911 AA.  
AC Q09056;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transferrin-binding protein 1 precursor.  
GN TBPL.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;  
RX MEDLINE=93345825; PubMed=8344530;  
RA Legrain M., Mazairin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,  
RA Jacobs E., Schryvers A.B.;  
RT \*Cloning and characterization of Neisseria meningitidis genes  
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.\*;  
RL Gene 130:73-80(1993).  
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR  
CC TRANSFERRIN UTILIZATION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- INDUCTION: By iron starvation.  
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z15130; CAA78833.1;  
CC PIR: JN0821; JN0821.  
CC InterPro: IPR000531; TonB\_boxC.  
CC Pfam: PF00593; TonB\_dep\_Rec.1.  
CC PROSITE: PS00430; TONB\_DEPENDENT\_REC.1; 1.  
CC PROSITE: PS01156; TONB\_DEPENDENT\_REC.2; 1.  
CC Outer membrane; Receptor; Signal; TonB box.  
KW SIGNAL 1 24  
FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.  
FT SITE 38 45 TONB\_BOX.  
FT SITE 894 911 TONB\_C-TERMINAL\_BOX.  
SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E6 CRC64;

Query Match 7.4%; Score 76; DB 1; Length 911;  
Best Local Similarity 26.4%; Pred. No. 22;  
Matches 60; Conservative 20; Mismatches 81; Indels 66; Gaps 13;  
QY 2 KKGSVIVGRINLSGDTAYAAQ-----TRGEGCQETSQ-----TGRDKNQ- 43  
Db 50 KRDNVETGLGLVKVTADTLSKEQVLDIRDLTRYDPLGLAVVEQGRGASSGYSIRGMDKNRV 109  
QY 44 ---VEGEVQIVSTAAQTFLATCINGVCWTVYHGAGRTIASPKGPVLOM-YTNVDKDLVG 99  
Db 110 SLTVDGLAQIOSYTAQAAL-----GGTET-AGSSGAINIEIENYKAVEIS 154  
QY 100 WPAQGSRLTPCTCGSSDL-----YLVTRHADVIPVRR-----GDSRGSLLSP 144  
Db 155 -----KGSNSVEQ---GSGALAGSVAFTKTADDVIGEGRWGIOSKATYSGKNRGLTQS- 206  
QY 145 RPISYLKSSGG--PLICPAGHAGVIFRA-AVCTRGVAKAVDFIPVE 188  
Db 207 ---TALAGRIGGAELLIHTGRRAGEIRAHEDAGRGVQSFNRLVPVE 250

Search completed: August 30, 2003, 19:13:45  
Job time : 10.7567 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run On: August 30, 2003, 19:00:22 : Search time 37.5921 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKSGSVVIGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL\_23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mmc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_todent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	910.5	88.2	4040	12 Q91FH8	Q91fh8 mucosal dis
2	901.5	87.4	2436	12 Q81756	Q81756 hepatitis c
3	901.5	87.4	3011	12 Q91FE5	Q91fe5 hepatitis c
4	900.5	87.3	3011	12 Q03463	Q03463 hepatitis c
5	896.5	86.9	3011	12 Q36579	Q36579 hepatitis c
6	896	86.8	181	12 Q91RR8	Q91rr8 hepatitis c
7	896	86.8	181	12 Q91RT5	Q91rt5 hepatitis c
8	894	86.6	181	12 Q91RR5	Q91rr5 hepatitis c
9	893	86.5	181	12 Q81LR3	Q81lr3 hepatitis c
10	893	86.5	181	12 Q91RS1	Q91rs1 hepatitis c
11	893	86.5	181	12 Q91R08	Q91r08 hepatitis c
12	893	86.5	181	12 Q91RT1	Q91rt1 hepatitis c
13	892.5	86.5	3011	12 Q91LS8	Q91ls8 hepatitis c
14	891	86.3	181	12 Q91RR6	Q91rr6 hepatitis c
15	891	86.3	181	12 Q91RS9	Q91rs9 hepatitis c
16	890	86.2	181	12 Q91RR2	Q91rr2 hepatitis c

17	890	86.2	181	12 Q91RS3	Q91rs3 hepatitis c
18	889.5	86.2	3011	12 Q9DIT6	Q9dit6 hepatitis c
19	889.5	86.2	3011	12 Q36608	Q36608 hepatitis c
20	889.5	86.2	3015	12 Q9PWX5	Q9pwx5 hepatitis c
21	889.5	86.2	3015	12 Q9PW09	Q9pwu9 hepatitis c
22	889	86.1	181	12 Q91RT4	Q91rt4 hepatitis c
23	889	86.1	181	12 Q91RS8	Q91rs8 hepatitis c
24	889	86.1	181	12 Q91RT3	Q91rt3 hepatitis c
25	889	86.1	181	12 Q91RS5	Q91rs5 hepatitis c
26	889	86.1	181	12 Q91RS7	Q91rs7 hepatitis c
27	889	86.1	181	12 Q91RT0	Q91rt0 hepatitis c
28	887	85.9	181	12 Q91RS4	Q91rs4 hepatitis c
29	886	85.9	181	12 Q91RT6	Q91rt6 hepatitis c
30	885	85.8	181	12 Q91RT9	Q91rt9 hepatitis c
31	884	85.7	181	12 Q91RR4	Q91rr4 hepatitis c
32	884	85.7	181	12 Q91RR9	Q91rr9 hepatitis c
33	884	85.7	181	12 Q91RR0	Q91rr0 hepatitis c
34	883.5	85.6	3011	12 Q36609	Q36609 hepatitis c
35	882	85.5	181	12 Q91RR7	Q91rr7 hepatitis c
36	881	85.4	181	12 Q91RT2	Q91rt2 hepatitis c
37	881	85.4	181	12 Q91RR1	Q91rr1 hepatitis c
38	881	85.4	181	12 Q91RQ9	Q91rq9 hepatitis c
39	881	85.4	181	12 Q91RS2	Q91rs2 hepatitis c
40	879	85.2	181	12 Q91RS6	Q91rs6 hepatitis c
41	878	85.1	181	12 Q91RT7	Q91rt7 hepatitis c
42	877	85.0	3011	12 Q36610	Q36610 hepatitis c
43	876	84.9	181	12 Q91RS0	Q91rs0 hepatitis c
44	876	84.9	181	12 Q91RT8	Q91rt8 hepatitis c
45	869.5	84.3	3010	12 Q9J3G9	Q9j3g9 hepatitis c

## ALIGNMENTS

### RESULT 1

ID	Q91FH8	PRELIMINARY;	PRT;	4040 AA.
AC	Q91FH8;			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Genome polyprotein.			
OS	Mucosal disease virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;			
OC	Pestivirus.			
OC	NCBI_TaxID=11099;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=20323484; PubMed=10864644;			
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,			
RA	Donis R.O., Hong Z., Lau J.Y.;			
RT	"Generation and characterization of a hepatitis C virus NS3 protease-			
RT	dependent bovine viral diarrhea virus.;"			
RL	J Virol. 74:6339-6347(2000).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lai V.C.H., Hong Z.;			
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
EMBL	AF268278; AAF82566.1; -			
HSSP	P26663; LJXP.			
MEROPS	S31.001; -			
DR	InterPro: IPR000280; CDvir_endptsep80.			
DR	InterPro: IPR001410; DEAD.			
DR	InterPro: IPR001409; HCV_NS3.			
DR	InterPro: IPR002166; HCV_RdRp.			
DR	InterPro: IPR001650; Helicase_C.			
DR	InterPro: IPR001005; Myb_DNA_binding.			
DR	InterPro: IPR001568; RNase_T2.			
DR	InterPro: IPR007095; RNA_pol_DS_PS.			
DR	InterPro: IPR007094; RNA_pol_PSVir.			
DR	Pfam: PF02907; HCV_NS3; 1.			
DR	Pfam: PF00271; helicase_C; 1.			
DR	Pfam: PF00998; Viral_RdRp; 1.			

```
DR PRINTS: PR00729; CDVENDOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
DR PROSITE: PS00531; RNASE_T2.2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
SO SEQUENCE 4040 AA; 453073 MW; ADE87791D055B9DC CRC64;

Query Match      88.2%; Score 910.5; DB 12; Length 4040;
Best Local Similarity 91.3%; Pred. No. 4.5e-83;
Matches 178; Conservative 5; Mismatches 9; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGEEGCOETSGTGRDKNOVEGEVQIVSTAAOTFLAT 61
DB 10 GSVVIVGRIVLSGSGSITACQAQOTRGLLGCKITSLTGRDKNOVEGEVQIVSTATOTFLAT 69
QY 62 CINGCVTVYHGAGTRTASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCTCGSSDLYL 121
DB 70 CINGCVTVYHGAGTRTASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCTCGSSDLYL 129
QY 122 VTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAKA 181
DB 130 VTRHANVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAVCTRGVAKA 189
QY 182 VDFIPVESLETTMRS 196
DB 190 VDFIPVENLETTMRS 204

RESULT 2
Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_Ns1.
DR InterPro; IPR002518; HCV_Ns2.
DR InterPro; IPR004109; HCV_Ns3.
DR InterPro; IPR000745; HCV_Ns4a.
DR InterPro; IPR001490; HCV_Ns4b.
DR InterPro; IPR002868; HCV_Ns5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01560; HCV_Ns1; 1.
DR Pfam; PF01538; HCV_Ns2; 1.
DR Pfam; PF02907; HCV_Ns3; 1.
DR Pfam; PF01006; HCV_Ns4a; 1.
DR Pfam; PF01001; HCV_Ns4b; 1.
DR Pfam; PF01506; HCV_Ns5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_Ns1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.

DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW RNA-directed RNA polymerase; Transferase; Polyprotein;
FT NON_TER 1 2436 2436
SO SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match      87.4%; Score 901.5; DB 12; Length 2436;
Best Local Similarity 85.8%; Pred. No. 1.9e-82;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

QY 3 KGSVWIVGRIN-----LSGDTAYAQOTRGEEGCOETSGTGRDKNOVEGEVQIVST 53
DB 555 RRGREILLGPADGMVSKGWELLAPITAYAQOTRGLLGCIITSLTGRDKNOVEGEVQIVST 614
QY 54 AAQTFLATCINGCVTVYHGAGTRTASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCT 113
DB 615 AAQTFLATCINGCVTVYHGAGTRTASPKGPVIQMTYNDKDLVGPAPQGSRLTPTCT 674
QY 114 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAV 173
DB 675 CGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPISYLKSGSGGPLLCPAGHAGVGFRAAV 734
QY 174 CTRGVAKAVDFIPVESLETTMRS 197
DB 735 CTRGVAKAVDFIPVENLETTMRS 758

RESULT 3
Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; AF271632; AAF81759.1; -.
DR HSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_Ns1.
DR InterPro; IPR002531; HCV_Ns2.
DR InterPro; IPR004109; HCV_Ns3.
DR InterPro; IPR000745; HCV_Ns4a.
DR InterPro; IPR001490; HCV_Ns4b.
DR InterPro; IPR002868; HCV_Ns5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_Ns1; 1.
```

```

DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC86AE58 CRC64;

Query Match 87.4%; Score 901.5; DB 12; Length 3011;
Best Local Similarity 85.8%; Pred. No. 2.5e-82;
Matches 175; Conservative 9; Mismatches 11; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDKNQVEGEVIVST 53
DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVIVST 1064

QY 54 AAQTFLATCINGVCVTVYHGAGTRTIAISPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 113
DB 1065 AAQTFLATCINGVCVTVYHGAGTRTIAISPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSSGGPLLPAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSSGGPLLPAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
Q03463 ID Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxId=11103;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;

```

```

RT HCV Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.*;
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchauspe G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.*;
RL Virology 200:246-255(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D10749; BAA01582.1; -.
DR HSP; P27958; IHEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

Query Match 87.3%; Score 900.5; DB 12; Length 3011;
Best Local Similarity 85.8%; Pred. No. 3.2e-82;
Matches 175; Conservative 8; Mismatches 12; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDKNQVEGEVIVST 53
DB 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVIVST 1064

QY 54 AAQTFLATCINGVCVTVYHGAGTRTIAISPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 113
DB 1065 AAQTFLATCINGVCVTVYHGAGTRTIAISPKGPVIQMTYNDKLVGWPAPQGSRLTPTCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSSGGPLLPAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSSGGPLLPAGHAVGIFRAAV 1184

```

```

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5
O36579 ID O36579 PRELIMINARY: PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSP; P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_Core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PS_vir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NSI; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NSI; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;

Query Match 86.98; Score 896.5; DB 12; Length 3011;
Best Local Similarity 84.8; Pred. No. 8.2e-82;
Matches 173; Conservative 10; Mismatches 12; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAAQTREGGCGQETSOTGRDKNQVEGEVIVST 53
DB 1005 RRGCEILLGADGMVSKGWRLLAPITAYAAQTREGGCGQETSOTGRDKNQVEGEVIVST 1064

RESULT 6
Q91RR8 ID Q91RR8 PRELIMINARY: PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Meyers D.L.;
RT "Genetic Diversity and Response to IFN of the NS3 Protease Gene from
RT Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369235; AAK54560.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
DR NON_TER 1 181
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D9186299B7C35 CRC64;

Query Match 86.88; Score 896; DB 12; Length 181;
Best Local Similarity 96.68; Pred. No. 2.5e-83;
Matches 172; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTREGGCGQETSOTGRDKNQVEGEVIVSTAAQTFLATCINGVCTVYHGAGT 78
DB 4 TAYAAQTREGGCGQETSOTGRDKNQVEGEVIVSTAAQTFLATCINGVCTVYHGAGT 63

QY 79 IASPKGPVQMTNVVDKDLVGVPAQPGSRSLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVQMTNVVDKDLVGVPAQPGSRSLTPTCGSSDLYLVTRHADVIPVRRGDSR 123

QY 139 GSLLSPRIPSYLKGSSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRIPSYLKGSSGGPLLCPCAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 7
Q91RT5 ID Q91RT5 PRELIMINARY: PRT; 181 AA.
AC Q91RT5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL.4;

```



```

RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369218; AAK54543.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 86.8%; Score 896; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 2.5e-83;
Matches 172; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAOQTRGEGCOETSQTRGKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAOQTRGLGCIITSLTGRKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

RESULT 8
QY1RR5
ID QY1RR5 PRELIMINARY; PRT; 181 AA.
AC QY1RR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.30;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369238; AAK54563.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match 86.6%; Score 894; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 4e-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAOQTRGEGCOETSQTRGKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAOQTRGLGCIITSLTGRKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

```

```

RESULT 9
QY1RR3
ID QY1RR3 PRELIMINARY; PRT; 181 AA.
AC QY1RR3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.4B;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369240; AAK54565.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19115 MW; 5D85F88AD7AC1A11 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.1e-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAOQTRGEGCOETSQTRGKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 78
DB 4 TAYAOQTRGLGCIITSLTGRKNOVEGEVQIVSTAAQTFLATCINGVCWTVYHGAGTRT 63
QY 79 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVIQMTYNDKDLVGNPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 181

RESULT 10
QY1RS1
ID QY1RS1 PRELIMINARY; PRT; 181 AA.
AC QY1RS1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.K;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF369232; AAK54557.1; -.
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3.1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19114 MW; ABB9085B3ABA4E26 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 5.1e-83;

```

Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTGREGGCGTSGTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 78  
 RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from  
 Clinical Strains of the Hepatitis C Virus.";  
 DB 4 TAYAAQTGRLGCGIITSLTGDRKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 63  
 QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138  
 DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123  
 QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196  
 DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

## RESULT 11

Q91R08  
 ID Q91R08 PRELIMINARY; PRT: 181 AA.  
 AC Q91R08;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE NS3 protease (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Pt.52;  
 RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;  
 RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from  
 Clinical Strains of the Hepatitis C Virus.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF369245; AAK54570.1; -;  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 KW Protease.  
 FT NON\_TER 1 1  
 FT NON\_TER 181 181  
 FT SEQUENCE 181 AA; 19144 MW; C0C91F1E2EEB0B32 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;  
 Best Local Similarity 96.1%; Pred. No. 5.le-83;  
 Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTGREGGCGTSGTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 78  
 DB 4 TAYAAQTGRLGCGIITSLTGDRKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 63  
 QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138  
 DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123  
 QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196  
 DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

## RESULT 12

Q91RT1  
 ID Q91RT1 PRELIMINARY; PRT: 181 AA.  
 AC Q91RT1;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE NS3 protease (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-Pt.161;  
 RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;  
 RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from  
 Clinical Strains of the Hepatitis C Virus.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF369222; AAK54547.1; -;  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 KW Protease.  
 FT NON\_TER 1 1  
 FT NON\_TER 181 181  
 FT SEQUENCE 181 AA; 19114 MW; ABB90B5B3ABA4E26 CRC64;

Query Match 86.5%; Score 893; DB 12; Length 181;  
 Best Local Similarity 96.1%; Pred. No. 5.le-83;  
 Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTGREGGCGTSGTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 78  
 DB 4 TAYAAQTGRLGCGIITSLTGDRKNQVEGEVQIVSTAAQTFLATCINGVCTVYHGAGT 63  
 QY 79 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138  
 DB 64 IASPKGPVIQMYTNVDKLVGWPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123  
 QY 139 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTMRS 196  
 DB 124 GSLLSPRISYLKSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTMRS 181

RESULT 13  
 Q9ELS8  
 ID Q9ELS8 PRELIMINARY; PRT: 3011 AA.  
 AC Q9ELS8;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Genome polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-colonel;  
 RA Desai S.M., Devare S., Yamaguchi J.;  
 RT "Hepatitis C Virus.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 PROTEIN C AND MRNA (BY SIMILARITY).  
 CC EMBL: AF290978; AAG02099.1; -;  
 DR HSP: P27958; IHEI  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.

```
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR Pfam: PF0186062; HCV_NS1; 1.
DR Pfam: PF00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match 86.5%; Score 892.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 2.1e-81;
Matches 172; Conservative 11; Mismatches 12; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAAQTRGEGCOETSTQGRDKNOVEGEVQIVST 53
DB 1005 RRGQELLGPADGMVSKGWRLLAPITAYAAQTRGLLGIITSLTGDRKNOVEGEVQIVST 1064
QY 54 AAQTFELATCINGVCVTVYHGAGTRTIAIPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCT 113
DB 1065 ATOTELATCINGVCVTVYHGAGTRTIAIPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCT 1124
QY 114 CGSSDLVLYTRHADVIVPVRRCDSGSLSPRISYLGSSGGPPLCPAGHAGVIFRAAV 173
DB 1125 CGSSDLVLYTRHADVIVPVRRCDSGSLSPRISYLGSSGGPPLCPAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETMRSP 1208

RESULT 14
Q91RR6 PRELIMINARY; PRT; 181 AA.
AC Q91RR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.3T;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369237; AAK54562.1;
DR InterPro: IP0004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19101 MW; 614ADA8B0F33CCAF CRC64;

Query Match 86.3%; Score 891; DB 12; Length 181;
Best Local Similarity 95.5%; Pred. No. 8.2e-83;
Matches 170; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCOETSTQGRDKNOVEGEVQIVSTAAQTFELATCINGVCVTVYHGAGTRT 78
DB 4 TAYAAQTRGLLGIITSLTGDRKNOVEGEVQIVSTAAQTFELATCINGVCVTVYHGAGTRT 63
QY 79 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 138
DB 64 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 123
QY 139 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRSP 196
DB 124 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRSP 181

Search completed: August 30, 2003, 19:18:19
Job time : 38.5921 secs
```

```
QY 79 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 138
DB 64 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 123
QY 139 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRSP 196
DB 124 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRSP 181

RESULT 15
Q91RS9 PRELIMINARY; PRT; 181 AA.
AC Q91RS9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pt.174;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
Clinical Strains of the Hepatitis C Virus.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369224; AAK54549.1;
DR InterPro: IP0004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19131 MW; 8BD7FC2769DBD635 CRC64;

Query Match 86.3%; Score 891; DB 12; Length 181;
Best Local Similarity 96.1%; Pred. No. 8.2e-83;
Matches 171; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGCOETSTQGRDKNOVEGEVQIVSTAAQTFELATCINGVCVTVYHGAGTRT 78
DB 4 TAYAAQTRGLLGIITSLTGDRKNOVEGEVQIVSTAAQTFELATCINGVCVTVYHGAGTRT 63
QY 79 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 138
DB 64 IASPKGPVIQMTNVKDLVGVWPAPOGSRSLTPTCTCGSSDLVLYTRHADVIVPVRRCDSR 123
QY 139 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETMRSP 196
DB 124 GSLLSPRPISYLGSSGGPPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETMRSP 181
```





A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A:Reference number: A40244; MUID:92230206; PMID:1314449  
A:Accession: A40244  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: GB:M84754  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructu  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 84.1%; Score 867.5; DB 1; Length 3010;  
Best Local Similarity 79.9%; Pred. No. 2.6e-71;  
Matches 163; Conservative 18; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVGEVQIVST 53  
Db 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLFGCIITSLTGRDNQVGEVQIVST 1064

QY 54 ATOTFLATCINGCVTVYHGAGTRTASPKGPVTOMYTNVDKLVGWAQPGSRLTPCT 113  
Db 1065 ATOSFLATCINGCVTVYHGAGSKTLAGPKGPTOMYTNVDQDLVGHAPGARSLLTPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPGSHAGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 5  
A45573  
genome polyprotein - hepatitis C virus (strain JT)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: A45573  
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,  
Virus Res. 23, 39-53, 1992  
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s  
A:Reference number: A45573; MUID:92295714; PMID:1318627  
A:Accession: A45573  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3010 <TAN>  
A:Cross-references: GB:D11168; GB:D01171; MID:g221612; PIDN:BAA01943.1; PID:g221613  
A:Experimental source: HCV-JT  
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBIP:106207)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: DEXH motif  
F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 83.1%; Score 857.5; DB 1; Length 3010;  
Best Local Similarity 78.4%; Pred. No. 2.2e-70;  
Matches 160; Conservative 20; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVGEVQIVST 53  
Db 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLGCIITSLTGRDNQVGEVQIVST 1064

QY 54 ATOTFLATCINGCVTVYHGAGTRTASPKGPVTOMYTNVDKLVGWAQPGSRLTPCT 113  
Db 1065 ATOSFLATCINGCVTVYHGAGSKTLAGPKGPTOMYTNVDQDLVGHAPGARSLLTPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPGSHAGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 6  
GNWVTC  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C:Accession: A38465  
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, I  
J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from hur  
A:Reference number: A38465; MUID:91140698; PMID:1847440  
A:Accession: A38465  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <TAK>  
A:Cross-references: EMBL:M58335; MID:g329770; PIDN:AAA72945.1; PID:g329771  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 82.7%; Score 853.5; DB 1; Length 3010;  
Best Local Similarity 77.9%; Pred. No. 5.1e-70;  
Matches 159; Conservative 21; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGECCOETSOTGRDNQVGEVQIVST 53  
Db 1005 RRGEILLGPADSLRGWRLAPITAYAAQOTRGLGCIITSLTGRDNQVGEVQIVST 1064

QY 54 ATOTFLATCINGCVTVYHGAGTRTASPKGPVTOMYTNVDKLVGWAQPGSRLTPCT 113  
Db 1065 ATOSFLATCINGCVTVYHGAGSKTLAGPKGPTOMYTNVDQDLVGHAPGARSLLTPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLKSGSGGGLLCPGSHAGVIFRAAV 1184





F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS>  
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,203

Query Match      65.6%; Score 677; DB 1; Length 3033;  
Best Local Similarity    68.7%; Pred. No. 9.8e-54;  
Matches     123; Conservative    26; Mismatches    30; Indels       0; Gaps       0;

Qy    19 TAYAAQTGREGCQETSOTGRDKNOVEGEVQIVSTATOTFLATCINGVCMTVIYHAGTTRI 78  
Db                  : ||||| : ||::|| | ||::|| | ||||| :  
1034 TAVAAQTGLLGTIIVVMTGRTKTEAQAGEIOVLSTVTSOSELCTISGVLTWTVYHGAKNT 1093

Qy    79 IASPKGPVTOMTYNDKDLYGWAOFGSRLSPCTCGSSDYLTLVRHADLVIPVRRGDNR 138  
Db                  : ||||| : ||::|| | ||::|| | ||||| :  
1094 LAGSRGPVTQMYSABEDLVGWPSPPGTSLSEPTCCGAVDLYLTRNADVIPARRRGDKR 1153

Qy    139 GSLLSPRPITLYLKSGGGPLCPACGHAVCIFRAA VCTRCAVKADVDFIPVESLETTMRSP 197  
Db                  : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

1154 GALLSPRJLTKLGSSGPGVPCLPRHAVGFVFAAACVSRCVAKSIDIFPVELDIVTRSP 1212

RESULT 11  
GNMWJ8

genome polypotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstruc  
protein NS4); nonstructural protein NS4B; nonstructural protein NS5  
C:Species: hepatitis C virus  
C>Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
Accession: A40250; PQ0397; PQ0559  
Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda,  
Virolgy 188, 331-341, 1992  
A>Title: Full-length sequence of a hepatitis C virus genome having poor homology to  
A:Reference number: A40250; MID:92230232; PMID:1314459  
Accession: A40250  
Molecule type: genomic RNA  
Residues: 1-3033 <OK>  
Cross-references: GB:D10988; GB:D01221; MID:g221608; PIDN:BAA01761.1; PID:g221609  
Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,  
J. Gen. Virol. 73, 1131-1141, 1992  
Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
Reference number: PQ0393; MID:92268871; PMID:1316939  
Accession: PQ0397  
Molecule type: genomic RNA  
Residues: 2678-2754 <CH>  
Cross-references: DDBJ:D10134  
Experimental source: isolate E-bl2  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
Title: Distribution of plural HCV types in Japan.  
Reference number: PQ0554; MID:92068204; PMID:1720309  
Accession: PQ0559  
Molecule type: mRNA  
Residues: 2678-2729 <KAT>  
Cross-references: GB:D10562; GB:D90518; MID:g221523; PIDN:BAA01418.1; PID:g221524  
Superfamily: Hepatitis C virus genome polyproteins  
Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F;1-115/Product: capsid protein C #status predicted <CP>  
F;116-191/Product: major envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <ME>  
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1011-1619/Product: hepacivirin #status predicted <NS3>  
F;1316-1321/Region: nucleotide-binding motif A (P-loop)  
F;1320-1323/Region: DEHX motif  
F;1620-1866/Product: nonstructural protein NS4 #status predicted <NS4>  
F;1867-2017/Product: nonstructural protein NS4B #status predicted <NS4B>  
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS>  
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,203

Query Match      65.4%; Score 675; DB 1; Length 3033;  
Best Local Similarity    69.3%; Pred. No. 1.5e-53;  
Matches     124; Conservative    24; Mismatches    31; Indels       0; Gaps       0;







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cn2.1/USPTO.spool/US09965594/runat\_29082003\_151919\_28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.ecgn\_1.1.14686.e\_runat\_29082003\_151919\_28310 -NCPU=6 -ICPU=3  
-NO\_MMALP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.scs.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pin.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	930.5	90.2	12734	6	AR179057 Sequence
2	910.5	88.2	12734	14	AF268278 Pestiviru
3	901.5	87.4	5360	6	AR118686 Sequence
4	901.5	87.4	5360	6	I06434 Sequence 48
5	901.5	87.4	5360	6	I09328 Sequence 8
6	901.5	87.4	6785	6	AR118692 Sequence
7	901.5	87.4	6785	6	I06440 Sequence 54
8	901.5	87.4	6785	6	I09329 Sequence 10
9	901.5	87.4	7310	6	AR118696 Sequence
10	901.5	87.4	7310	6	I09331 Sequence 15
11	901.5	87.4	7310	14	HPCPOLYP M32084 Hepatitis C
12	901.5	87.4	8316	6	AR118703 Sequence
13	901.5	87.4	8987	6	AR118728 Sequence
14	901.5	87.4	9185	6	AR118722 Sequence
15	901.5	87.4	9185	6	AR118723 Sequence
16	901.5	87.4	9185	6	BD091382 Sequence
17	901.5	87.4	9185	6	I08294 Sequence 1
18	901.5	87.4	9379	6	AR166930 Sequence
19	901.5	87.4	9379	6	AR301300 Sequence
20	901.5	87.4	9401	6	AR176483 Sequence
21	901.5	87.4	9401	6	BD080334 Hepatitis
22	901.5	87.4	9401	6	E66593 Hepatitis C
23	901.5	87.4	9401	6	I71894 Sequence 9
24	901.5	87.4	9401	6	I81885 Sequence 9
25	901.5	87.4	9401	14	HPCPLYPRE M62321 Hepatitis C
26	901.5	87.4	9609	12	AF387805 Synthetic
27	901.5	87.4	9609	12	AF387808 Synthetic
28	901.5	87.4	9618	14	AF271632 Hepatitis
29	901.5	87.4	9646	12	AF387806 Synthetic
30	901.5	87.4	9693	12	AF387807 Synthetic
31	900.5	87.3	9502	6	E08263 gRNA of Hep
32	900.5	87.3	9502	6	E08264 cDNA of Hep
33	900.5	87.3	9502	14	HPCRCJ1 DI0749 Hepatitis C
34	899	87.1	2058	6	AX395309 Sequence
35	899	87.1	2058	6	AX454818 Sequence
36	899	87.1	8157	6	AR127810 Sequence
37	899	87.1	8157	6	BD081911 Hepatitis
38	898.5	87.1	9424	14	AF511948 Hepatitis
39	897	86.9	1932	6	AR127809 Sequence
40	897	86.9	1932	6	BD081910 Hepatitis
41	896.5	86.9	9646	6	AR110828 Sequence
42	896.5	86.9	9646	6	BD069982 Functiona
43	896.5	86.9	9646	14	AF009606 Hepatitis
44	896.5	86.9	12980	6	AR110831 Sequence
45	896.5	86.9	12980	6	BD069985 Functiona

#### ALIGNMENTS

RESULT 1

AR179057  
LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2000  
DEFINITION Sequence 1 from patent US 6326137.  
ACCESSION AR179057  
VERSION AR179057.1 GI:20220612  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12734)  
AUTHORS Hong,Z., Lai,V.C.H. and Lau,J.Y.N.  
TITLE Hepatitis C virus protease-dependent chimeric pestivirus  
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;  
FEATURES  
source location/Qualifiers  
1..12734 /organism="unknown"  
BASE COUNT 4032 a 2604 c 3295 g 2803 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5 27e-65 Length: 12734  
Score: 930.50 Matches: 181  
Percent Similarity: 94.87% Conservative: 4  
Best Local Similarity: 92.82% Mismatches: 7  
Query Match: 90.16% Indels: 3  
DB: Gaps: 1

US-09-965-594-14 (1-197) x AR179057 (1-12734)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 413 GGTAGTGTGTATTGTGTGTAGAAATGTTTATCTGCTAGTGGTAGTATCATCGCGCTAC 472  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
|||||  
Db 473 GCCCAGCAGCAGAGGCGCTCTTAGGGTGTAGATCATCCGCTCTGACTGGCGGGGACAAA 532  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61  
|||||  
Db 533 AACCAAGTGGAGGTGAGGTCCAGATCGTCTCAACTGCTACCCAAACCTTCTGSCAACG 592  
QY 62 CysIleAsnGlyValCysTTPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 593 TGCATCAATGGGGTATGCTGGACTGTCTACACGGGGCGGAACGAGGACCATCGCATCA 652  
QY 82 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPro 101  
|||||  
Db 653 CCCAAGGCTCTGTCATCCAGATGATACCAATGTGGACCAAGACTTGTGGCTGGCCC 712  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 713 GCTCTCAAGGTTCCTGCTCAATTGACACCTGTCACCTGCGGCTCTCGACCTTTACCTG 772  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 773 GTTAGAGCAGCGCAGGTCATTCCTCGCGCGGGAGGTGATAGCAGGGGTAGCCTG 832  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 833 CTTTTCGCCCGCGCCATTCTCTACTAAAAGGCTCTCGGGGGGTCCGCTGTGTGTGCCCC 892  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
Db 893 GCGGNCACCGCGTGGGCTATTACGGCGCGGGTGTGCACCCGTGGAGTGGCCAAAGCG 952  
QY 182 ValAspPheIleProValIgluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 953 GTGGACTTTATCCCTGTGGAAACCTAGACAAACCATGAGATCC 997

RESULT 2  
AF268278  
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000  
DEFINITION Pestivirus type 1, complete genome.

3'UTR  
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others  
ORIGIN  
VALFGVYGVYQALSKRRVPMITDIYTTIEDORLEDTIHLQYAPNAIKTGDGTETELKELAS  
GDVEKINGALISDYAAGGLEFVKSOAQEKITAPLRFKNAEAGASGVQKFDISLTENKEE  
IIRYGLWGTATFLYKSLAARLGHETAFATLVKLWAFGESVSUHVQKQALVDLVYVY  
MKNPSPGDSFETOQEGRRVVASLFI SALATYTYKTYNHNLSKVLESPALAYLPYATSA  
LKMFPTPRLBSVLSSTTYTKTILSRKSGDLGLGTGISAANEIISQNPVSVGISVM  
LGVGTAAHNAIIESSQRTLLMKVFKNFLDQAATDELVKENPEKILMALFEAVQTI  
GNPLRIYHLUGVYKYGWEAKESERTAGNCLFTLMPFAFELGLMDSGKIRNLSGN  
YIIDLIIYGLHQRNGLKMKVLGWAAPAFSCDWTSPDERILRPLDNLVRYETPCOGY  
EMAFKNVGGKLTIKVSPESGFLCNRNRPGRPVNRYTKYDDNLEIKIPVAKLEGQVE  
HYKGVTAKTIDYSGKMLATDKAEVHEGVITLAKRYTGVGPNGAYLGDENPHRALV  
ERCATITKNTVQFLMKKCACTYDLTISNLRLELHVRNNLEKEIPTAIVTWL  
AYTFVNDVGTIKPVLIGERVIPDPVDINDLOPEVQDITSEVGTIIIGRETLMITGTP  
VLEKVEPDSNDNSVKIGDEGNPGPGIQTHTLTEEIHNRDARFIMILGSRNSIS  
NRAKTARNILNYTGNDRPRETRIDMAARMLVALRDVPELSMWDFKGTFLDREALE  
ALSUGPKPKQVTKKAVRNLEIQQKQVEIPNPFASDDPVFEVALANDKXYLVLDGVE  
VKDQAKALGATDQTRILIKVGSRTYAMKLSWFLOASNKOMSLTLPFELLRCPPAT  
KSNKGHMASAYQLAOGNNEPLGCGVHLGTIPARRVKIHPYEAULKLDIIEEEKAPR  
VKDTVIREHNNKILIFOGNLTNMLNPKGLSSELODREGRKKNYHNOIKIMSS  
AGIRLEKLPTRAQDTKTPEHAIROKIDKSENQNPENLHKLLEIHTIAQTLKHT  
YGEVTFQLEAGVNRKGAAGLEKKNIGEVLDSEKHLVQOLVRLDKAGRKIKYETAI  
PKNEKRDVSDMOAGDLVVKRPVIOYPEAKTRLAITKVYNNVQOOPVJIPYEGK  
TPLFNIFDKVKRKEWDSNEPVAVSFDTKAMDQVTSKDLQELICEIOKYIYKKEWRFI  
DTIDHNTVEPVITADGEVIRNCRGSGQPDISAGNSMLNVLTMWAFCESTGVYK  
SFNRVARIHVGDDGDFLITBEGGLAFANKGMQILHEAGRPQKITEGEMKAVAFRED  
IEFCSHTPPVPRSDNTSSHMAGRDTAVILSKMATRLDSSGERGTTAYEKAVAFSL  
MYSNPLVRRICLILVSOQPTDPSKHAITYYKGDPIGAYKDVIGRNLSELKRTGFPEK  
LANLNLISLTGLTWTHTSKRIIQDCVAIKERGNMLVNADRLISSTGLHYIPDKGF  
TLOGKHVEQLQLATEINPNVGVCTERYKLGPIVNLRLRLKILLMTAVGVSS"  
12509..12734  
Alignment Scores:  
Pred. No.: 2,55e-63 Length: 12734  
Score: 910.50 Matches: 178  
Percent Similarity: 93.85% Conservative: 5  
Best Local Similarity: 91.28% Mismatches: 9  
Query Match: 88.23% Indels: 3  
DB: 14 Gaps: 1  
US-09-965-594-14 (1-197) x AF268278 (1-12734)  
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 413 GGTAGTGTGTATTGTGTGAGAAATGTTTATCTGCTAGTGTAGTATCACGGCGTCG 472  
Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
Db 473 GCCCAGCAGCAGGAGGCGCTCTAGGGTGTGAAGATCACCACTCTGCTGGCGGGGACAA 532  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61  
Db 533 AACCAAGTGGAGGTGAGGTCCAGATCGTGCACTGCTACCCAAACCTTCTTGGCAACG 592  
Qy 62 CysIleAsnGlyValCysTrpThrValTyRHISGlyAlaGlyThrArgThrIleAlaSer 81  
Db 593 TGCATCAATGGGGTATGCTGGACTGTCTACACCGGGCGGCAACGAGGACCATCGCATCA 652  
Qy 82 ProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 101  
Db 653 CCCAAGGCTCTGTGATCCAGATGATACCAATGTGGACCAAGACCTTGTGGGTGCGCC 712  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 713 GCTCCTCAAGGTTCCTCGCTCATTCGACACCTGACCTGCGGCTCTCTCGGACCTTTACCTG 772  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 773 GTTACAGGCACGCAACGCTATTCCTCCGCGCGCGGCGAGGTGATAGCAGGGGTAGCCTG 832  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 833 CTTTGGCCCCGGCCCATTTCTTACATAAAGGCTCCTCTGGGGGTCCGCTGTGTGTGCCCC 892

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
Db 893 GCGGGACACCGCGTGGGCTATTTCAGGCGCGGGTGTGCACCGGTGGAGTGGCCAAAGGCG 952  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
Db 953 GTGGACTTATCCCTGTGGAGAACCTAGAGACAACGACGAGATCC 997  
RESULT 3  
ARL18686  
LOCUS ARL18686 5360 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 53 from patent US 6150087.  
ACCESSION ARL18686  
VERSION ARL18686.1 GI:14100596  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Chien,D.Y.  
TITLE NANBV diagnostics and vaccines  
JOURNAL Patent: US 6150087-A 53 21-NOV-2000;  
FEATURES Location/Qualifiers  
source  
1..5360  
/organism="unknown"  
BASE COUNT 1060 a 1623 c 1532 g 1145 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.32e-63 Length: 5360  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 6 Gaps: 1  
US-09-965-594-14 (1-197) x ARL18686 (1-5360)  
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 867 CGCAGGGCGGGAGATGCTGCTGGGCGCAGCGGATGGTAATGCTTCCAAAGGGGTGGAGG 926  
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 927 TTGCTGCGCCCATCATCGGCGTACGCCAGCAGACACAAAGGGGCTCTAGGGTGCATAATC 986  
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 987 ACCAGGCTTAACGTGGCGGGACAAAAACCAAGTGGAGGGTGAGGTCCAGATTGTGCAACT 1046  
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1047 GCTGCCCAAACTTCTTGGCAAGCTGCATCAATGGGCTGTGTGGACTGTCTACCAAGG 1106  
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1107 GCCGGAACGAGGACCATCGCGTCACCAAGGTCCTCATCCAGATGATACCAATGTA 1166  
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1167 GACCAAGACCTTCTGGGCTGGCGCGTCCGCAAGGTAGCGCTCATTTGACACCTGCACT 1226  
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 1227 TGGCGCTCTCTGGACCTTTACCTGGTCACGAGCAGCGCATGTATCCCGGTGGCGCG 1286  
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 1287 CGGGGTGATAGCAGGGGACCGCTGCTGCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346  
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173  
Db 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173

Db 1347 TCGGGGGTCCGCTGTTGTGCCCCCGGGGCACGCCGTGGCATATTTAGGGCCGCGTG 1406  
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 1407 TGCACCGGTGGAGTGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466  
Qy 194 MetArgSerPro 197  
Db 1467 ATGAGGTCCCCG 1478  
RESULT 4  
LOCUS I06434 5360 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 48 from Patent EP 0318216.  
ACCESSION I06434  
VERSION I06434.1 GI:590311  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton, M., Choo, Q.-L. and Kuo, G.  
TITLE Nanbv diagnostics and vaccines  
JOURNAL Patent: EP 0318216-A1 48 31-MAY-1989;  
FEATURES  
Location/Qualifiers  
1..5360  
source /organism="unknown"  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,32e-63 Length: 5360  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: Gaps: 1  
US-09-965-594-14 (1-197) x I06434 (1-5360)  
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 867 CGCAGGGGCGGGAGATGCTCGGGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 926  
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 927 TTGCTGGCGCCATCAGCGGTAGCCAGCAGACAAGGGGCTCCTAGGTGCATAATC 986  
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53  
Db 987 ACCAGCCTAACTGCCCGGGACAAAACCAAGTGGAGGGTGAGTCCAGATTGTGTCAACT 1046  
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1047 GCTGCCCAAACTTCTGGCAACTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGG 1106  
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1107 GCCGGAACGAGGACCATCGCGTCACCAAGGGTCTCATCAGATGTATACCAATGTA 1166  
Qy 94 AspLysAspLeuValGlyTTPProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGTAGCGCTCATTTGACACCTCACT 1226  
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 1227 TCGGGCTCTCGGACCTTTACTGTGTACGAGGACGCCGATGTCTCCGTGCGCCGG 1286  
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 1287 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346  
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173

Db 1347 TCGGGGGTCCGCTGTTGTGCCCCCGGGGCACGCCGTGGCATATTTAGGGCCGCGTG 1406  
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 1407 TGCACCGGTGGAGTGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 1466  
Qy 194 MetArgSerPro 197  
Db 1467 ATGAGGTCCCCG 1478  
RESULT 5  
LOCUS I09328 5360 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 8 from Patent WO 8904669.  
ACCESSION I09328  
VERSION I09328.1 GI:587963  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5360)  
AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.  
JOURNAL Patent: WO 8904669-A 8 01-JUN-1989;  
FEATURES  
Location/Qualifiers  
1..5360  
source /organism="unknown"  
BASE COUNT 1061 a 1623 c 1533 g 1143 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,32e-63 Length: 5360  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: Gaps: 1  
US-09-965-594-14 (1-197) x I09328 (1-5360)  
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 867 CGCAGGGGCGGGAGATGCTCGGGCCAGCGCATGGAATGCTCCAAAGGGGTGGAGG 926  
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 927 TTGCTGGCGCCATCAGCGGTAGCCAGCAGACAAGGGGCTCCTAGGTGCATAATC 986  
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53  
Db 987 ACCAGCCTAACTGCCCGGGACAAAACCAAGTGGAGGGTGAGTCCAGATTGTGTCAACT 1046  
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1047 GCTGCCCAAACTTCTGGCAACTGCATCAATGGGGTGTCTGGACTGTCTACCAAGGG 1106  
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1107 GCCGGAACGAGGACCATCGCGTCACCAAGGGTCTCATCAGATGTATACCAATGTA 1166  
Qy 94 AspLysAspLeuValGlyTTPProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1167 GACCAAGACCTTGTGGGCTGGCCGCTCCGCAAGTAGCGCTCATTTGACACCTCACT 1226  
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 1227 TCGGGCTCTCGGACCTTTACTGTGTACGAGGACGCCGATGTCTCCGTGCGCCGG 1286  
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 1287 CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 1346  
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173

```
|||||
Db 1347 TCGGGGGTCCGCTGTTGTGCCCCGGGGCAGCCGTGGGCATATTTAGGCGCGGGTG 1406
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1407 TGCACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGAGAACCTTAGACAACC 1466
Qy 194 MetArgSerPro 197
Db 1467 ATGAGGTCCCGC 1478

RESULT 6
LOCUS AR118692 6785 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 65 from patent US 6150087.
ACCESSION AR118692
VERSION AR118692.1 GI:14100602
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Chien,D.Y.
TITLE NAMV diagnostics and vaccines
JOURNAL Patent: US 6150087-A 65 21-NOV-2000;
FEATURES
    Location/Qualifiers
        1..6785
            /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 6.86e-63 Length: 6785
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x AR118692 (1-6785)
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1140 CGCAGGCGCGGGAGATGCTCTCGGCCACGCCGGAATGTCTCCAAAGGGTGGAGG 1199
Qy 15 ---LeuSerGlyAspThrAlaValAlaGlnThrArgGlyGluGlyCysGlnGlu 33
Db 1200 TTGCTGGGCCCATCAGCGGTAGCGCCAGCAGACAGGGGCTCCTTAGGGTGCATATC 1259
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1260 ACCAGCCTAATCGCGCGGACAAAACCAAGTGGAGGTCAGGTCCAGATTGTCTCAACT 1319
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 1320 GCTGCCCAAACTTCTCGGCAACGTGATCAATGGGGTGTGCTGACTGTCTACACGGG 1379
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1380 GCCGGAACGAGGACCATCGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1439
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1440 GACCAAGACCTTGTGGGTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 1499
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCAGGAGCAGCCGATGTTCATTCCTGCGCGG 1559
Qy 134 ArgGlyAspSerArgGlySerLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 1560 CGGGGTGATACAGGGGCGAGCCTGCTGTGCCCCCGGCCCAATTTCTCTACTTGAAGGCTCC 1619
```

```
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 1620 TCGGGGGTCCGCTGTTGTGCCCCGGGGCAGCGGTGGGCATATTTAGGCGCGGGTG 1679
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 1680 TGCACCGGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGAGAACCTTAGACAACC 1739
Qy 194 MetArgSerPro 197
Db 1740 ATGAGGTCCCGC 1751

RESULT 7
LOCUS 106440 6785 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 54 from Patent EP 0318216.
ACCESSION 106440
VERSION 106440.1 GI:590312
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6785)
AUTHORS Houghton,M., Choo,Q.-L. and Kuo,G.
TITLE Nanbv diagnostics and vaccines
JOURNAL Patent: EP 0318216-A1 54 31-MAY-1989;
FEATURES
    Location/Qualifiers
        1..6785
            /organism="unknown"
BASE COUNT 1392 a 2050 c 1914 g 1429 t
ORIGIN

Alignment Scores:
Pred. No.: 6.86e-63 Length: 6785
Score: 901.50 Matches: 175
Percent Similarity: 90.20% Conservative: 9
Best Local Similarity: 85.78% Mismatches: 11
Query Match: 87.35% Indels: 9
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x 106440 (1-6785)
Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 1140 CGCAGGCGCGGGAGATGCTCTCGGCCACGCCGGAATGTCTCCAAAGGGTGGAGG 1199
Qy 15 ---LeuSerGlyAspThrAlaValAlaGlnThrArgGlyGluGlyCysGlnGlu 33
Db 1200 TTGCTGGGCCCATCAGCGGTAGCGCCAGCAGACAGGGGCTCCTTAGGGTGCATATC 1259
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 1260 ACCAGCCTAATCGCGCGGACAAAACCAAGTGGAGGTCAGGTCCAGATTGTCTCAACT 1319
Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 1320 GCTGCCCAAACTTCTCGGCAACGTGATCAATGGGGTGTGCTGACTGTCTACACGGG 1379
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93
Db 1380 GCCGGAACGAGGACCATCGCTCACCAAGGGTCTCTCATCCAGATGTATACCAATGTA 1439
Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 1440 GACCAAGACCTTGTGGGTGGCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 1499
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 1500 TCGGGCTCTCGGACCTTTACCTGGTCAGGAGCAGCCGATGTTCATTCCTGCGCGG 1559
Qy 134 ArgGlyAspSerArgGlySerLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 1560 CGGGGTGATACAGGGGCGAGCCTGCTGTGCCCCCGGCCCAATTTCTCTACTTGAAGGCTCC 1619
```

QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173  
Db 1620 TCGGGGGTCCGCTGTGTGTCGGCCGGGACGGCTGGCATATTTAGGGCCGGGTG 1679  
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193  
Db 1680 TGCACCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCTAGAGAAC 1739  
QY 194 MetArgSerPro 197  
Db 1740 ATGAGTCCCG 1751  
RESULT 8  
LOCUS I09329 6785 bp DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 10 from Patent WO 8904669.  
ACCESSION I09329  
VERSION I09329.1 GI:587964  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 6785)  
AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.  
JOURNAL Patent: WO 8904669-A 10 01-JUN-1989;  
FEATURES  
Location/Qualifiers  
1..6785  
Source /organism="unknown"  
BASE COUNT 1392 a 2050 c 1914 g 1429 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.86e-63 Length: 6785  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 6 Gaps: 1  
US-09-965-594-14 (1-197) x I09329 (1-6785)  
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 1140 CGCAGGGCCGGAGATACACTGCTGGCCACCGCATGGAAATGCTCCAGGGGTGGAGG 1199  
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33  
Db 1200 TTGCTGGGCCCATCATCAGCGCTAGCCAGACAGAGGGCCCTCCTAGGGTGCATAATC 1259  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 1260 ACCAGCTTAATGCGCGGACAAAACCAAGTGGAGGGTGAAGTCCAGATGTGTCAACT 1319  
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTyrThrValTyrHisGly 73  
Db 1320 GCTGCCCAACCTTCCTGGCAACGTGATCAATGGGGGTGCTGGACTGCTACACGGG 1379  
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1380 GCCGGAACGAGGACCATCGGCTCACCCAAAGGTCTCTCATCCAGATGTATACCAATGTA 1439  
QY 94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1440 GACCAAGACCTTGTGGGTGGCCGGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAT 1499  
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 133  
Db 1500 TGGCGCTCCTCGGACCTTACCTGGTCAGGAGCAGCCGATGTCTATCCCGTGGCCGG 1559  
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 1560 CGGGGTGATAGAGGGGACGCTGCTGCGCCCGGCCCATTTCTTCTACTTGAAGGCTCC 1619

QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173  
Db 1620 TCGGGGGTCCGCTGTGTGTCGGCCGGGACGGCTGGCATATTTAGGGCCGGGTG 1679  
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193  
Db 1680 TGCACCGTGGAGTGGCTAAGGGGTGGACTTTATCCCTGTGGAGAACCTAGAGAAC 1739  
QY 194 MetArgSerPro 197  
Db 1740 ATGAGTCCCG 1751  
RESULT 9  
LOCUS ARI18696 7310 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 74 from patent US 6150087.  
ACCESSION ARI18696  
VERSION ARI18696.1 GI:14100606  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 7310)  
AUTHORS Chien, D.Y.  
JOURNAL NABV diagnostics and vaccines  
Patent: US 6150087-A 74 21-NOV-2000;  
FEATURES  
Location/Qualifiers  
1..7310  
Source /organism="unknown"  
BASE COUNT 1495 a 2220 c 2056 g 1539 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.44e-63 Length: 7310  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 6 Gaps: 1  
US-09-965-594-14 (1-197) x ARI18696 (1-7310)  
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 1665 CGCAGGGCCGGAGATACACTGCTGGCCACCGCATGGAAATGCTCCAGGGGTGGAGG 1724  
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33  
Db 1725 TTGCTGGGCCCATCATCAGCGCTAGCCAGACAGAGGGCCCTCCTAGGGTGCATAATC 1784  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 1785 ACCAGCTTAATGCGCGGACAAAACCAAGTGGAGGGTGAAGTCCAGATGTGTCAACT 1844  
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTyrThrValTyrHisGly 73  
Db 1845 GCTGCCCAACCTTCCTGGCAACGTGATCAATGGGGGTGCTGGACTGCTACACGGG 1904  
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1905 GCCGGAACGAGGACCATCGGCTACCCAAAGGTCTCTCATCCAGATGTATACCAATGTA 1964  
QY 94 AspLysAspLeuValGlyTyrProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1965 GACCAAGACCTTGTGGGTGGCCGGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAT 2024  
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 133  
Db 2025 TGGCGCTCCTCGGACCTTACCTGGTCAGGAGCAGCCGATGTCTATCCCGTGGCCGG 2084  
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153



Db 2085 CGGGGTGATAGCAGGGCAGCGTCTGCTCGCCCGGCCCATTTCTACTTGAAGGCTCC 2144

Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
|||||  
Db 2145 TCGGGGGTCCCGTGTGTGCGCCCGGGCAGCGGTGGCATATTATTAGGCCCGGGTG 2204  
|||||

Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193  
|||||  
Db 2205 TGCACCGGTGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAAC 2264  
|||||

Qy 194 MetArgSerPro 197  
|||||  
Db 2265 ATGAGGTCCCG 2276  
|||||

RESULT 10

LOCUS I09331 7310 bp DNA linear PAT 02-DEC-1994

DEFINITION Sequence 15 from Patent WO 8904669.

ACCESSION I09331

VERSION I09331.1 GI:587966

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 7310)  
AUTHORS Houghton, M., Choo, Q.-K. and Kuo, G.  
JOURNAL Patent: WO 8904669-A 15 01-JUN-1989;

FEATURES  
Location/Qualifiers  
1..7310  
/organism="unknown"

BASE COUNT 1495 a 2218 c 2058 g 1539 t

ORIGIN

Alignment Scores:  
Pred. No.: 7 44e-63 Length: 7310  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 6 Gaps: 1

US-09-965-594-14 (1-197) x I09331 (1-7310)

Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
|||  
Db 1665 CGCAGGGCGGGAGATACTGCTCGGCCAGCGATGGAATGCTCTCAAGGGGTGGAGG 1724  
|||

Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33  
|||||

Db 1725 TTGCTGGCCCGCATCAGCGGTACCGCCAGCAGCAAGGGGCTCCTAGGTGCTAATC 1784  
|||||

Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
|||||

Db 1785 ACCAGCCTAACTGCGCGGCACAAAACCAAGTGGAGGTGAGTCCAGATTGTGTCAACT 1844  
|||||

Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleThrValThrHisGly 73  
|||||

Db 1845 GTGTGCCAAACCTCTCGGCAACGTGCATCAATGGGGGTGCTGACGTGTCTACACGGG 1904  
|||||

Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetIleThrAsnVal 93  
|||||

Db 1905 GCGGAGACGAGGACCATCGCTACCCAGGGTCTGTCTATCCAGTGTATACCANTGTA 1964  
|||||

Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
|||||

Db 1965 GACCAAGACCTTGTGGCTGGCCGCTCGCAAGTAGCGCTCATTGACACCGCTGCAC 2024  
|||||

Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
|||||

Db 2025 TCGCGTCTCGGACCTTTACCTGTGTCAGGAGCAGCGCGATGCTCATTCCTCGCGCGG 2084  
|||||

Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
|||||

Db 2085 CGGGGTGATAGCAGGGCAGCGTCTGCTCGCCCGGCCCATTTCTACTTGAAGGCTCC 2144

Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
|||||  
Db 2145 TCGGGGGTCCCGTGTGTGCGCCCGGGCAGCGGTGGCATATTATTAGGCCCGGGTG 2204  
|||||

Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193  
|||||  
Db 2205 TGCACCGGTGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGACACAAC 2264  
|||||

Qy 194 MetArgSerPro 197  
|||||  
Db 2265 ATGAGGTCCCG 2276  
|||||

RESULT 11

LOCUS HPCPOLYP 7310 bp ss-RNA linear VRL 02-AUG-1993

DEFINITION Hepatitis C virus polyprotein gene, partial cds.

ACCESSION M32084

VERSION M32084.1 GI:329875

KEYWORDS polyprotein.

SOURCE Hepatitis C virus

ORGANISM Hepatitis C virus

REFERENCE 1 (bases 1 to 7310)  
AUTHORS Choo, Q.-L., Richman, K. and Han, J.  
JOURNAL The nucleotide sequence of the Hepatitis C viral genome  
Unpublished (1990)  
COMMENT Original source text: Hepatitis C virus, cDNA to viral RNA, clones K9-1 through 15e, isolated from chimpanzee (individual 910) blood plasma.  
Draft entry and printed sequence for [1] kindly submitted by M. Houghton, 22-FEB-1990. Chiron Corporation, 4560 Horton Street, Emeryville CA 94608.

FEATURES  
Location/Qualifiers  
1..7310  
/organism="Hepatitis C virus"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:11103"  
<1..>7310  
/note="polyprotein"  
/codon\_start=3  
/protein\_id="AAA45677.1"  
/translation="GCPEERLASRLPTDFDQGWGPISYANGSGPQDPQPCWHPKPKPC  
GIVPAKSVCGPVYCFTPSPVYGTDRSGAPTSWGENDTDFVLNNTRPPLGNWPGC  
TWNSTGFTKCGCAGPCPVIGGAGNNTLHCPDFCRKHPDATYSCGSGPWITPRCLVD  
YPRLHYECTINYTIETKIRMYGVGVEHLEACNWTGERCDLEDRSELSPLLT  
TTQWVPCSFITLPAISGLHLHONIYDVQILYGVSSIASWALKWIFVYLLFLL  
ADARVCSUMMLLIISOAAALENIVLNAASLAGTHGLVSLVFFCFAMYLUKRWVP  
GAYTTYGMPLLLLLALPORAYALDTEVAASCGGVLVGLMALTLSPPYKRYISWC  
LWMLQYFLTRVEAQLHVMIPPLNVRGRDAVILMCVHPTLVFDITKLLAVFGPLW  
ILQALSLTPYFVRVQGLLRFCALARKMTGGHVQMVIIKLGALGTYYVYHLTPRD  
WANGRLDLAVAVEPVVSOMETKLITMGADTAACGDIINGLPVSAARRRETLIGPAD  
GMYSKWRLLAPITAYAOQTRGLLGGIITSLGRDNKNOVEGVQIVYSTAAQFLATCI  
NGVCVTYHAGTRTASRGPVIOMTINVDQLVGWPAQPSRSLTPCTCGSSDLX  
VTRHADVIPVRRRSGRGLSPRIYSLKSGSGGLPCPAGHVGIFPAACTRGVA  
KYVDLIPVENLETTMSPVFTDNSSPVVPSQFVAHLHAPTHAPTSKSTKFLDAAGQ  
KAVFLNVAATLFGFAMSKAHGIDPNIRGTVRTITGSPITTYSTYKFLADGCS  
GGAVDITICDECHSDATSLIGTGLVDOAGTAGARLVATATPPGTVVPHNTEE  
VALSTGEPPYCKAIPLEVIKGRHLIFCHSKKCKDELAALVALGINAVAYRGLD  
VSVPTSGVVATDALMTGTGDFSDVDICNTCTVTFSLDPTFTTETILPQD  
AVSKTRRGTRGKPGIYFVYFAPRPSGMSDFSVLCYCDAGCAWELTAPETVR  
LRAMMTPLGLPCQDHLFEWEGVTGLTHDAFLTSQKSGENLPYLVAYQATCAR  
AQAPPSWQWQMKLIRLAPTHGPTPLLYRLGAVONEITLTHPTVKYIMTQMSADLE  
VVTSTVYGGVLAALAAVCLSTGCYVIVGVVLSKPAIIPDREVLVREFDEBES  
QHLPYTEQGMLEAQFKQALGLQLQASROAEVIAFVQTNWKLTFEFAKHMWFCIS  
GIQYLAGSLTLPGNPAISLMFAATVTSPLTTSQTLFNLGGWVAQALAAPGATA  
FVGAGLAGAAIGSVGLGKVLIDILAGYAGAGALVAFKIMSGEVPSTEDLVNLLPAI  
LSPGALVGVVCAAILRRHVPGEGAVQNMNRLIAFASRGNRYSPHYVPESDAARV  
TAILSLITVQLLRHLRHLSSPECTPCSGSLRDLTDWHTICEVLSFKTLKALMPO  
LPGIPFVSCORGKGVWRVDGIMHTRCHGAEITGVKNGKTHRIVGPRICRNWSGTF

PINATYTGCTPLPAPNTFALWRYSAEYVEIROVGFHVTGTTDNLKCPQVPS  
PEFFTELDGVRUHRFAPCKPLREEVFRVGLHEYPVGSQPCPEPDVAULTSMT  
DPSHITAAGRLARSPSPVASSASOLSAPSKATCTANHDSPPDAELTEANLWR  
QMGNGNITRVESNKVWILSDPDLVAREDEREISVPAEILRKSRPAQALPVWARP  
YNPLVETWKKPDYEPVPHGCPPLPKSPVPVPPKRTVVLBESTLSALAEATR  
SFGSSSTGIDNTTISEPAPSCPDSDAESYSSMPLEPEGPDLDSDGSMSTV  
SSEANEDVCCSMYSWTGALVTPCAAEQKLPIINALNSLLRHNDLVYTTSRAC  
QKRVTFDRQLVSHYODVLKEYAAASKANLSSVEEACSLTPPHSAKSFYGG  
AKDVRCHARKAVHTNSVWKDLLENDVTPIDTINAKNEVFCVQPKGGRKPARLIV  
POLGVRCVKMAYLVDVTKPLAVMGSSVGFQYSPQORVEFLVOAKSKKTPMGFSYD  
TRCFDSTVETSDIRTEBALYOCDDLDQARVAIKSLTERLYVGGPLTNSRNGCYRR  
CRASGVLTSCGNTLITIKAKACRAGLQCTMLVCGDDLVVICESAGVEDRASL  
RAFTAMRYTSAPDGPPOPEYDELEITSCSNVSVAHGAGKRVYIITRODPTPLAR  
AAWETARHTPTVNSGTIMEFAPTLWARMILMTHFVSFLIARDQLEQALDEIYGACY  
STIEPLDLPPIIOLR\*

BASE COUNT 1495 a 2218 c 2058 g 1539 t  
ORIGIN

## Alignment Scores:

Pred. No.: 7,44e-63 Length: 7310  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 14 Gaps: 1

US-09-965-594-14 (1-197) x HPCPOLYP (1-7310)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 1665 CGCAGGCGCGGAGATATCTCGCGCCAGCGGATGGTATGTCACCAAGGGTGGAGG 1724  
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33  
Db 1725 TTGCTGGCGCCCATCATCGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGTCATAATC 1784  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 1785 ACCAGCTAACTGGCGCGGACAAAACCAAGTGGAGGTGGAGTGCAGATTGTGCAACT 1844  
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1845 GCTGCCCAAACTTCTGCAACGTCATCAATGGGTGTGCTGGACGTCTACCAAGGG 1904  
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1905 GCGGAACAGGAGGACCATCGCTACCAAGGTCCTGTCTCATCCAGATGTATACCAATGTA 1964  
QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlnGlySerArgSerLeuThrProCysThr 113  
Db 1965 GACCAAGACTGTGGCGTGGCGCGCTCCGCAAGGTAGCGCTCATTTGACACCCCTGCAC 2024  
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 2025 TGGCGCTCTCGGACCTTTACCTGGTCACGAGGACGCGGATGTCATTCCCGTGGCGCG 2084  
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 2085 CGGGGTGATAGCAGGGGACCGCTGTCTCGCGCGGCCCATTTCTCTACTTGAAGGCTCC 2144  
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
Db 2145 TCGGGGGTTCGCTGTGTGCGCGCGGACCGGTGGGATATTTAGGCGCGCGGTG 2204  
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 2205 TGCACCGGTGAGCTGAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 2264  
QY 194 MetArgSerPro 197  
Db 2265 ATGAGGTCCCGC 2276

RESULT 12

ARI18703  
LOCUS ARI18703 8316 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 88 from patent US 6150087.  
ACCESSION ARI18703  
VERSION ARI18703.1 GI:14100613  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 8316)  
AUTHORS Chien,D.Y.  
TITLE NAMBV diagnostics and vaccines  
JOURNAL Patent: US 6150087-A 88 21-NOV-2000;  
FEATURES Location/Qualifiers  
source 1..8316  
BASE COUNT 1671 a 2529 c 2345 g 1771 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 8,55e-63 Length: 8316  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 6 Gaps: 1  
US-09-965-594-14 (1-197) x ARI18703 (1-8316)  
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 2671 CGCAGGCGCGGAGATATCTCGCGCCAGCGGATGGTATGTCACCAAGGGTGGAGG 2730  
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluCysGlnGlu 33  
Db 2731 TTGCTGGCGCCCATCATCGCGGTACGCCAGCAGACAAAGGGGCTCTCTAGGTGTCATAATC 2790  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 2791 ACCAGCTAACTGGCGCGGACAAAACCAAGTGGAGGTGAGGTGTCAGATTGTGCAACT 2850  
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 2851 GCTGCCCAAACTTCTCGGCAACGTCATCAATGGGTGTGCTGGACTGTCTACCAAGGG 2910  
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 2911 GCGGAACAGGAGGACCATCGCTACCAAGGTCCTGTCTCATCCAGATGTATACCAATGTA 2970  
QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlnGlySerArgSerLeuThrProCysThr 113  
Db 2971 GACCAAGACTTGTGGCGTGGCGCGCTCCGCAAGGTAGCGCTCATTTGACACCCCTGCAC 3030  
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 3031 TGGCGCTCTCGGACCTTTACTGTGTACGAGGACACCGCATGTATCCGTGGCGCGG 3090  
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 3091 CGGGGTGATAGCAGGGGACGCTGTCTCGCGCGGCCCATTTCTCTACTTGAAGGCTCC 3150  
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
Db 3151 TCGGGGGTTCGCTGTGTGCGCGCGGACCGGTGGGATATTTAGGCGCGCGGTG 3210  
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 3211 TGCACCGGTGAGGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 3270  
QY 194 MetArgSerPro 197  
Db 3271 ATGAGGTCCCGC 3282

LOCUS	ARL18728	Sequence	137 from patent US 6150087.	DNA	linear	PAT 16-MAY-2001
DEFINITION	ARL18728	Sequence	137 from patent US 6150087.	DNA	linear	PAT 16-MAY-2001
ACCESSION	ARL18728	Sequence	137 from patent US 6150087.	DNA	linear	PAT 16-MAY-2001
VERSION	ARL18728.1	GI:14100638				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 8987)					
AUTHORS	Chien,D.Y.					
TITLE	NANBV diagnostics and vaccines					
JOURNAL	Patent: US 6150087-A 137 21-NOV-2000;					
FEATURES	Location/Qualifiers					
source	1..8987					
BASE COUNT	1807 a	2735 c	2547 g	1898 t		
ORIGIN	/organism="unknown"					
Alignment Scores:						
Pred. No.:	9.3e-63	Length:	8987			
Score:	901.50	Matches:	175			
Percent Similarity:	90.20%	Conservative:	9			
Best Local Similarity:	85.78%	Mismatches:	11			
Query Match:	87.35%	Indels:	9			
DB:	6	Gaps:	1			
US-09-965-594-14 (1-197) x ARL18728 (1-8987)						
QY	3	LysLysGlySerValValIleValGlyArgIleAsn	-----14			
DB	3013	CGCAGGGCGGGAGATATCTGCTGGGCCAGCGATGGAATGGTCTCCAAGGGTGGAGG	3072			
QY	15	---LeuSerGlyAspThrAlaTyAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu	33			
DB	3073	TTGCTGGCGCCCATCATCGCGCTACGCCAGCAGACAGAGGGGCTCTTAGGGTGATATC	3132			
QY	34	ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr	53			
DB	3133	ACCAGGCTTAACCTGGCGGGACAAAACCAAGTGGAGGGTGAGTCCAGATTGTGTCAACT	3192			
QY	54	AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyHisGly	73			
DB	3193	GCTGCCCAACCTTCTGCCACGTGCATCATGGGGTGCTGGACTGTCTACCCAGGG	3252			
QY	74	AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyThrAsnVal	93			
DB	3253	GCGGGAACGAGGACCATCGCTGCCAAGGGTCTGTGTATCCAGATATATACCAATGA	3312			
QY	94	AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr	113			
DB	3313	GACCAAGACTTGTGGGCTGGCGGCTCCGCAAGGTAGCCGCTCAITGACACCTGTCACT	3372			
QY	114	CysGlySerSerAspLeuTyLeuValThrArgHisAlaAspValIleProValArgArg	133			
DB	3373	TGCGGCTCTCGGACCTTATCTGTGTGACAGGACGCGCGATGTCATCCCGTGCGCGG	3432			
QY	134	ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer	153			
DB	3433	CGGGGTGATAGCAGGGGACGCTGCTGTCGCCCGGCCCATTTCTACTTGAAGGCTCC	3492			
QY	154	SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal	173			
DB	3493	TGCGGGGGTCCGCTGTGTGTCGCCCGGGGACGCGCGTGGGCATATTAGGGCCCGG	3552			
QY	174	CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuThrThr	193			
DB	3553	TGCACCGGTGGACTGCTAGGGGGTGGACTTATCCCTGTGGAGAACCTTAGAGCAACC	3612			
QY	194	MetArgSerPro 197				
DB	3613	ATGAGGTCCCGG 3624				



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-14

Perfect score: 1032

Sequence: 1 MKKXGVSIVIGRLNSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model  
-O=/cpn2.1/usptc03pool/US09965594/runat.29082003.151918.28302/app\_query.fasta.1.2872  
-DB=N\_Geneseq.19Jun03 -QEXT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594\_8CGN\_1.1\_1412-brunat.29082003.151918.28302 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq.19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	1032	100.0	594	21	AAA73330	Hepatitis C virus	
2	1015	98.4	594	21	AAA73331	Hepatitis C virus	
3	998	96.7	588	21	AAA73329	Hepatitis C virus	
4	993	96.2	594	21	AAA73335	Hepatitis C virus	
5	985	95.4	594	21	AAA73332	Hepatitis C virus	
6	973	94.3	594	21	AAA73333	Hepatitis C virus	
7	963	93.3	594	21	AAA73334	Hepatitis C virus	
8	959	92.9	588	21	AAA73328	Hepatitis C virus	
9	930.5	90.2	12734	24	ABA95615	Chimeric BVDV/HCV	
10	921.5	89.3	612	25	ABX15706	Anti-viral synthet	
11	901.5	87.4	5300	10	AAN92097	Combined open read	
12	901.5	87.4	5360	10	AAN90327	Hepatitis C virus	
13	901.5	87.4	6905	10	AAN92103	Combined open read	
14	901.5	87.4	7310	10	AAN92106	Combined open read	
15	901.5	87.4	7310	10	AAN90336	Composite hepatiti	
16	901.5	87.4	7310	16	AAQ98221	Hepatitis C virus	
17	901.5	87.4	8316	21	AAA75296	cDNA sequence comp	
18	901.5	87.4	9133	20	AAZ07656	Nucleotide sequenc	
19	901.5	87.4	9185	11	AAO05956	Sense strand of th	
20	901.5	87.4	9185	12	AAQ10566	Hepatitis C virus	
21	901.5	87.4	9185	21	AAA75297	Sense strand of HC	
22	901.5	87.4	9400	13	AAQ21744	Compiled HCV cDNA.	
23	901.5	87.4	9401	17	AAT12710	Hepatitis C virus	
24	901.5	87.4	9401	18	AAT99981	HCV polyprotein co	
25	901.5	87.4	9401	19	AAV09989	Hepatitis C virus	
26	901.5	87.4	9401	24	AAD35043	Hepatitis C virus	
27	900.5	87.3	9502	15	AAQ74770	Hepatitis C virus	
28	899	87.1	549	21	AAA70344	Hepatitis C virus	
29	899	87.1	2058	24	AAZ29795	Hepatitis C virus	
30	899	87.1	2058	24	ABK15344	HCV-1 NS3/4a mutan	
31	899	87.1	2058	25	ABX14410	Hepatitis C virus	
32	899	87.1	8145	20	AAZ32359	DNA encoding HCV-1	
33	898.5	87.1	9185	20	AAZ26737	Plasmid pBT-BS(+)	
34	898.5	87.1	9185	20	AAZ00459	Nucleotide sequenc	
35	897.5	87.0	8316	11	AAQ05955	Hepatitis C virus	
36	897	86.9	1933	20	AAZ32358	Hepatitis C virus	
37	896.5	86.9	9646	19	AAV59361	HCV NS3 DNA. Hepa	
38	896.5	86.9	9646	24	ABK87285	cDNA encoding hepa	
39	896.5	86.9	12980	19	AAV59364	Hepatitis C virus	
40	896.5	86.9	12980	24	ABK87286	Hepatitis C virus	
41	896.5	86.9	16622	21	AAZ36212	Nucleotide sequenc	
42	895	86.7	2061	24	AAD34500	Hepatitis C virus	
43	895	86.7	2061	24	AAD31767	Hepatitis C virus	
44	894.5	86.7	1998	20	AAZ80355	HCV NS4A-NS3 compl	
45	892.5	86.5	9365	24	AAD25518	Hepatitis C virus	

ALIGNMENTS

RESULT 1  
AAA73330  
ID AAA73330 standard; DNA; 594 BP.  
XX  
XX  
AC AAA73330;  
XX  
XX  
19-DEC-2000 (first entry)  
DT  
XX  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.  
XX  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; muten; ds.  
XX  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
XX  
FH Key Location/Qualifiers

FT CDS 1..594  
 FT /\*tag= a  
 FT /product= "NS4A-NS3 fusion protein #3"  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15221.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 26; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0-1 variant.  
 XX  
 SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other:

Alignment Scores:  
 Pred. No.: 2,42e-85 Length: 594  
 Score: 1032.00 Matches: 197  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73330 (1-594)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAATAAAAGAGTCCGTTGTTATCGTCGGCGGTATCAACCTGTCGGTGACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnThrSerGlnThrGlyArgAsp 40  
 DB 61 TACGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAACCTCCAGACGGTGTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60  
 DB 121 AAACACAGGTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTCCTGGCT 180  
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTGCATCAGCGTGTGTGTCAGCGTTTACACCGGTGCTGTACCGTACCATCGCT 240  
 QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 DB 241 TCCCGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGTGG 300  
 QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerAspLeuTyr 120

DB 301 CGGGTCGGCAGGGTTCGGTTCCTGCACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGGTTACCGCTCAGCGTGCAGCTTATCCCGGTTCCTCGTGTGAGTCCCGTGGTTC 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGCGCTCCGATCTCTACCTGAAGGTTCCTCGGGTGTCCGCTGCTGTC 480  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
 DB 481 CGGGTGGTCAAGCTGTGGTATCTTCGCTGCTGTGTTCACCCGCTGGTGTCTAA 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCTGTGACTTCATCCCGGTTCGAATCCCTGGAAACCAACCATGCTGCTCCCG 591  
 RESULT 2  
 AAA73331  
 ID AAA73331 standard; DNA; 594 BP.  
 XX  
 AC AAA73331;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT CDS 1..594  
 FT /\*tag= a  
 FT /product= "NS4A-NS3 fusion protein #4"  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15222.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 26; Fig 14; 66pp; English.  
 XX  
 CC The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0-1 variant.







```

PD 13-JUL-2000.
XX
XX
XX PF 06-JAN-2000; 2000WO-US00345.
XX
XX PR 08-JAN-1999; 99US-0115271.
XX
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX DR WPI: 2000-465976/40.
XX
XX DR P-PSDB; AAB15223.
XX
XX PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX PS Claim 26; Fig 15; 66pp; English.
XX
XX CC The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
XX SQ Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:
Pred. No.: 4.72e-81 Length: 594
Score: 985.00 Matches: 191
Percent Similarity: 96.95% Conservative: 0
Best Local Similarity: 96.95% Mismatches: 6
Query Match: 95.45% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73332 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGATCCGTTGTTATCGTGGCGGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 40
DB 61 TACGCTCAGCAGACTCGAGGTGAGGAGGGTGGCAAGAAACCTCCAGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAAAAACGAGTTGAAGTGAAGTTTCAGATGTTTCCACCGCTACCCAGACTTCCTCGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGGACGGTTTACACCGTGTCTGACCGGTACCATCGCT 240
QY 81 SerProLysGlyProValIleGlnMetTyThrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGAAGGTCGGGTACCAGATGTACCAACAGTGTGACAAAGACCTGGTGGTGG 300
QY 101 ProIleProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
DB 301 CAGGCTCCGACAGGGTCCCGTTCCTGACCCCGTGCACCTCGCGTTCCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCCGTCAGCTGAGTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420

```

---

```

QY 141 LeuLeuSerProArgProIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 421 CTGCTGTCGCCGCGTCCGATCTCTACCTGAAGGTTCTCCGCGTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAAGCTGTTGGTATCTCTCCGCTGCTGCTGCTTCCACCGTGGTGTCTCTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCCATCGGTGCCCG 591

RESULT 6
AAA73333
ID AAA73333 standard; DNA: 594 BP.
XX
XX AC AAA73333;
XX
XX DT 19-DEC-2000 (first entry)
XX
XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
XX
XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; muten; ds.
XX
XX OS Hepatitis C virus.
XX
XX FH Key Location/Qualifiers
XX CDS 1..594
XX FT /tag= a
XX FT /product= "NS4A-NS3 fusion protein #6"
XX
XX PN WO200040707-A1.
XX
XX PD 13-JUL-2000.
XX
XX PF 06-JAN-2000; 2000WO-US00345.
XX
XX PR 08-JAN-1999; 99US-0115271.
XX
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX DR WPI: 2000-465976/40.
XX
XX DR P-PSDB; AAB15224.
XX
XX PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX PS Claim 26; Fig 16; 66pp; English.
XX
XX CC The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX
XX SQ Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:
Pred. No.: 5.88e-80 Length: 594
Score: 973.00 Matches: 188

```

```

Percent Similarity: 96.45% Conservative: 2
Best Local Similarity: 95.43% Mismatches: 7
Query Match: 94.28% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73333 (1-594)

QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAATAAAAGGATCCGTTGTTATCGCGCCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCAGCAGACGTCGAGGTGAGCAGGTTGCCAGAGACCTCCACACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTTGAAGGTGAAGTTGAGTTCAGATCGTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTCTCTGTGGACCGTTTACACGGTGTGTTACCGGTACCATCGCT 240
QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGGAAAGGTCGCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTCCGTTCCCTGACCGCGTGACCTGCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGCTGTCCCGGTCGATCCTCCGTTATCCCGTTTCCGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGTCGATCCTCCGTTGAAAGGTTCCTCCGCGGTGCTGCTGCTGCTGCTGCTGCT 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTTGTTATCCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTACTTATCCCGTTGAATCCCTCGAAACACCATCGTTCCTCCCG 591

RESULT 7
AAA73334
ID AAA73334 standard; DNR; 594 BP.
XX
AC AAA73334;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutin; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
Key Location/Qualifiers
FT CDS 1..594
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #7"
XX
XX WO200040707-A1.
XX
PN 13-JUL-2000.
XX
PD 06-JAN-2000; 2000WO-0500345.
XX
PF

```

```

XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
XX
DR P-PSDB; AAB15225.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 17; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-7 variant.
XX
SQ Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

Alignment Scores:
Pred. No.: 4.82e-79 Length: 594
Score: 963.00 Matches: 187
Percent Similarity: 95.94% Conservativeness: 2
Best Local Similarity: 94.92% Mismatches: 8
Query Match: 93.31% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-14 (1-197) x AAA73334 (1-594)
QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAATAAAAGGATCCGTTGTTATCGCGCCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCAGCAGACGTCGAGGTGAGCAGGTTGCCAGAGACCTCCACACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTTGAAGGTGAAGTTGAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGGACCGTTTACACGGTGTGTTACCGGTACCATCGCT 240
QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGGAAAGGTCGCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 300
QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTTCGCTTCCCTGACCGCGTGACCTGCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGCTGTCCCGGTCGATCCTCCGTTATCCCGTTTCCGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGTCGATCCTCCGTTGAAAGGTTCCTCCGCGGTGCTGCTGCTGCTGCTGCTGCT 480

```

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaValCysThrArgGlyValAlaLys 180  
 DDb 481 CCGCGTGGTCACCGTGTGTATCTCCGTCGCTGCTTTCCACCCGCGTGTGTCTAA 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DDb 541 GCTGTTGACTTCATCCCGGTTGAATCCTGGAACACCACCATGCGTTCCCGG 591

RESULT 8  
 AAA73328  
 ID AAA73328 standard; DNA: 588 BP.  
 AC AAA73328;  
 XX  
 XX 19-DEC-2000 (first entry)  
 DT  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 1..588  
 FT /\*tag= a  
 FT /product= "NS3-NS4A fusion protein"  
 XX  
 PN W0200040707-A1.  
 XX  
 XX 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX  
 XX 08-JAN-1999; 99US-0115271.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15212.  
 XX  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT .  
 XX  
 XX Disclosure; Fig 10; 66pp; English.  
 PS  
 XX  
 CC The present sequence is the coding sequence for a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.  
 XX  
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

## Alignment Scores:

Pred. No.: 1,1e-78 Length: 588  
 Score: 959.00 Matches: 188  
 Percent Similarity: 95.94% Conservative: 1  
 Best Local Similarity: 95.43% Mismatches: 6  
 Query Match: 92.93% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-14 (1-197) x AAA73328 (1-588)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DDb 1 ATGAAAAAAGGTTCCGTTGTTATCGTCGGCCGTATAGTACTGAACGGT-----GCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyAlaGsp 40  
 DDb 55 TAGCGTCAGCAGACTCGAGGTCGTGGGTGGTCATCATCCTCCCTGACCGGTCTGTGAC 114  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60  
 DDb 115 AAAAACCCAGGTTGAAGGTGAAGTTCAGATGCTTTCCACCGCTGCTCAGACCTTCCTGGCT 174  
 QY 61 ThrCysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DDb 175 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCGGTACCATCGCT 234  
 QY 81 SerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
 DDb 235 TCCCGGAAAGTCCGGTTATCCAGATGTACACCAACGTTGCACAAAGACCTGTTGGTTGG 294  
 QY 101 ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DDb 295 CCGGCTCCGCAGGTTCCCGTTCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DDb 355 CTGTTTACCGCTCAGCGTCAGCTTATCCCGGTTGCTGCTGCTGCTGCTGCTGCTGCTCC 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DDb 415 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCTCCCGTGGTCCGCTGCTGCTGTC 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
 DDb 475 CCGGCTGTCACGCTGTTGTTATCTTCCGTGCTGCTTTGCACCCGCTGCTGTTGCTAAA 534  
 QY 181 AlavalAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DDb 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCACCATGCGTTCCCGG 585

## RESULT 9

ABA95615  
 ID ABA95615 standard; DNA: 12734 BP.  
 XX  
 AC ABA95615;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Chimeric BVDV/HCV NS3-wt sequence.  
 XX  
 KW Pestivirus; Npro; protease; NS3; screening; ds.  
 XX  
 OS Chimeric - Bovine viral diarrhea virus.  
 OS Chimeric - Hepatitis C virus.  
 XX  
 PN US6326137-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 PF 25-JUN-1999; 99US-0344456.  
 XX  
 PR 25-JUN-1999; 99US-0344456.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Hong Z, Lai VCH, Lau JYN;  
 XX  
 DR WPI; 2002-121103/16.  
 XX  
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3  
PT protease gene, useful for in vivo screening of compounds which inhibit  
PT HCV infection  
XX  
XX Example 2; Columns 17-28; 20pp; English.  
PS  
XX The present invention relates to a nucleic acid construct encoding a  
CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
CC each junction site recognised by the Npro protease is replaced with a  
CC junction site recognised by the HCV NS3 protease. The construct is useful  
CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
CC protease, where screening may be in cell culture or in an animal model.  
CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
CC virus)/HCV NS3-wt, which was used to illustrate the present invention.  
XX  
SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:  
Pred. No.: 1-94e-74 Length: 12734  
Score: 930.50 Matches: 181  
Percent Similarity: 94.87% Conservative: 4  
Best Local Similarity: 92.82% Mismatches: 7  
Query Match: 90.16% Indels: 3  
DB: 24 Gaps: 1

US-09-965-594-14 (1-197) x ABA95615 (1-12734)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
DB 413 GSTAGTGTGTTTATGTTAGTAAATTTTATCTGTGTAGTGTATCATCGGGGTAC 472  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
DB 473 GCCCAGCAGCAGAGCCCTCTAGGGTGAAGATCACCAGTCTGACTGGCGGGACAAA 532  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThr 61  
DB 533 AACCAAGTGGAGGGTCCAGATCGTGTCAACTGCTACCCAAACCTTCCTTGGCAACG 592  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyValGlyThrArgThrIleAlaSer 81  
DB 593 TGCATCAATGGGTATGCTGGACTGTCTACACGGGGCCGGAACGAGGACCATCGCATCA 652  
QY 82 ProGlyGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpPro 101  
DB 653 CCCAAGGGTCTGTCTCCAGATGTATACCAATGTGACCAAGACCTGTGGGCTGGCCC 712  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 713 GCTCCTCAAGGTTCCCGCTCATTCACACCTCGACCTCGCGGCTCTCGGACCTTTACCTG 772  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
DB 773 GTACGAGGACCGCCGACGTATCCCGTCCGCGGCGAGGTGATACGAGGGTAGCCTG 832  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
DB 833 CTTTGGCCCGGCCCCATTCTCTACCTAAAGGCTCTCTCGGGGGGTGCGTGTGTGCCCC 892  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
DB 893 GCGGAGACCGCGTGGGCCCTATTTCAGGGCCCGGTGTGCACCGGTGGAGTGGCAAGGCG 952  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGACACCAACCATGATGCC 997  
RESULT 10  
ID ABX15706  
XX US-09-965-594-14 (1-197) x ABX15706 (1-612)

AC ABX15706;  
XX 28-MAR-2003 (first entry)  
XX  
XX Anti-viral synthetic prototoxophore associated DNA sequence.  
DE  
XX Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;  
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;  
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.  
XX  
OS Unidentified.  
XX WO200287500-A2.  
XX 07-NOV-2002.  
XX 26-APR-2002; 2002WO-US13223.  
XX 27-APR-2001; 2001US-286893P.  
XX (NEWB-) NEWBIOTICS INC.  
XX Cathers BE, Neuteboom STC, Shepard HM;  
XX WPI: 2003-167102/16.  
XX  
PT Novel synthetic viral prototoxophore for treating viral infections, has  
PT toxin moiety incorporated into substrate domain specific for viral  
PT enzyme, bound and modified by viral enzyme to get converted into  
PT toxophore  
XX  
XX Example 1; Page 62; 66pp; English.

This invention relates to a novel synthetic viral prototoxophore comprising a toxin moiety operatively incorporated into a substrate domain specific for a viral enzyme. This prototoxophore may be bound and modified by the viral enzyme thus converting it to a toxophore. Also disclosed in the invention is a method for enhancing the anti-viral effect of an antiviral agent, this method comprises contacting a cell, infected with a virus or is susceptible to infection, with a prototoxophore. The invention further comprises an assay to identify anti-viral agents, comprising contacting an infected cell with a candidate agent and comparing the ability of the agent to inhibit the growth or infectivity of the virus in the cell. The prototoxophores of the invention may have virucide or antitumour activity. The prototoxophores of the invention may be useful for reducing or inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte, nerve cell, connective tissue cell, muscle cell or hepatocyte), which infected with a virus or is susceptible to infection with a virus, with an effective amount of the prototoxophore. The cells are cell lines adapted to long term continuous culture or isolated from a subject. The prototoxophore is also useful for ameliorating the severity of a viral infection in a subject, where the virus is selected from human immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and hepatitis virus, by administering an effective amount of the prototoxophore to the subject. The prototoxophores of the invention are also useful for treating tumours. The present sequence represents an antiviral prototoxophore associated DNA sequence, this sequence is described as a recombinant NS3/NS4 fusion protein in example 1 of the invention although it is clearly not a protein sequence.

XX Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

Alignment Scores:  
Pred. No.: 3.07e-75 Length: 612  
Score: 921.50 Matches: 181  
Percent Similarity: 94.36% Conservative: 3  
Best Local Similarity: 92.82% Mismatches: 8  
Query Match: 89.29% Indels: 3  
DB: 25 Gaps: 1

(CHIR ) CHIRON CORP.  
Houghton M, Choo QL, Kuo G;  
WPI: 1989-159274/22.  
P-PSDB; AAP92041.  
Purified hepatitis C virus  
- and associated nucleic acids and polypeptide(s)  
Claim 3; Figure 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.  
It is a double-stranded nucleotide sequence of the open reading frame (ORF) (tag a) extending through clones 141, 11b, 7f, 7e, 8h 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c of hepatitis C virus (HCV) cDNA. In creating the composite sequence the following heterogeneities were considered. Clone 33c contains a sequence of 800 base pairs which overlaps the cDNAs in clones 40b and 37c. In clone 33c, as well as in 5 other overlapping clones, nucleotide 1789 is a G. However, in clone 37b the corresponding nucleotide is an A. This heterogeneity may have important ramifications for protein folding. Nucleotide #2 in clone 8h is a T which may represent a cloning artifact because the corresponding residue in clone 7e and in 3 other overlapping clones is an A. Therefore the residue in this position is designated as an A. The 3'-terminal nucleotide in clone 8f is represented as a T than a G because the correspondign residue in clone 33f and in 2 other overlapping clones is a T. The 3' terminal sequence of clone 33f is represented as ATTC, as is found in the corresponding sequence in clone 33g and in 2 other overlapping clones, rather than as TTGC, as is found in clone 33f. Residue #4 in clone 33g is designated an A rather than a T because the corresponding residue in clone 33f and 2 other overlapping clones is an A. The 3'-terminus of clone 141 is depicted as TA rather than AA because the corresponding dinucleotide in clone 11b and 3 other clones is TA. Potential cloning artifacts have been omitted and instead the corresponding sequences in non-5'-terminal regions of multiple overlapping clones are shown. AAN92097 could be used as a source of oligomeric DNA hybridisation probes to detect the presence of HCV nucleic acids in samples. The polypeptide(s) it encodes could be used as immuno- assay reagents and vaccines and to generate antibodies useful in diagnosis and passive immunotherapy for HCV infection/non-A, non-B hepatitis.  
(Updated on 25-MAR-2003 to correct PR field.)  
(Updated on 25-MAR-2003 to correct PI field.)  
XX Sequence 5300 BP; 1047 A; 1606 G; 1515 G; 1130 T; 2 other;  
Alignment Scores:  
Pred. No.: 2,93e-72 Length: 5300  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
DB: 10 Gaps: 1  
US-09-965-594-14 (1-197) x AAN92097 (1-5300)  
QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db CGCAGGGCGCGGAGATACTGCTCGGCCACCGCATGAATGGTCTCCAAGGGGTGGAGG 926  
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db TTTCTGGGCCCCATCATCGGCGTATGCGCCACACACAGGGGCCCTCTAGGTGCATAATC 986  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53  
Db ACCAGCCTAACTGCGCGGGACAAANAAACCAAGTGGAGGTGAGGTCCAGATGTGTCAACT 1046  
QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysIleAsnGlyValCysIleHisGly 73  
Db GCTGCCAAACCTTCCTGGCAACGTGCATCAATGGGGGTGTGCTGGAGCTGTCTACCAGGG 1106

QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
 DB 1107 GCCGGAACAGGACCATCGTCACCAAGGTCCTGTTCATCCAGATGTATACCAATGTA 1166  
 QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
 DB 1167 GACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCAC 1226  
 QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
 DB 1227 TCGGGCTCTCGGACCTTACCTGTGTCAGAGCAGCGCGATGTATTCCTGCGCGG 1286  
 QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
 DB 1287 CGGGGTGATAGCAGGGCAGCCCTGTGTGCGCCCGGCCCATTTCTACTTGAAGGCTCC 1346  
 QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173  
 DB 1347 TCGGGGGGTCCGCTGTGTGCCCCCGGGGCGACGCGGTGGGCATATTTAGGGCCCGGGT 1406  
 QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
 DB 1407 TGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 1466  
 QY 194 MetArgSerPro 197  
 DB 1467 ATGAGGTCCCG 1478

## RESULT 12

AAN90327  
 ID AAN90327 standard; cDNA; 5360 BP.

AC AAN90327;

XX 25-MAR-2003 (updated)

DT 11-NOV-1989 (first entry)

XX Hepatitis C virus composite probe.

DE Hepatitis C virus; composite cDNA; probe; vaccine.

XX Pan troglodytes.

XX Key Location/Qualifiers

FT CDS 3..5360

FT /\*tag- a

XX GB2212511-A.

XX 26-JUL-1989.

XX 18-NOV-1988; 88GB-0027024.

XX 18-NOV-1987; 87US-0122714.

XX 30-DEC-1987; 87US-0139886.

XX 26-FEB-1988; 88US-0161072.

XX 26-OCT-1988; 88US-0263584.

XX (CHIR ) CHIRON CORPORATION.

XX Houghton M, Choo QL, Kuo G;

XX WPI; 1989-215054/30.

XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,

XX polypeptide(s) and antibodies for diagnosis, prevention and treatment

XX Of infection.

XX Disclosure; Fig. 26; 174pp; English.

XX The sequence shows the composite cDNA sequence derived from the aligned

XX hepatitis C virus (HCV) cDNA's in clones 14i, 11b, 7f, 7e, 8h, 33c, 40b,

CC 37b, 35, 36, 8i, 32, 33b, 25c, 14c, 8f, 33f, 33g and 39c. The cDNA

CC encodes antigens which react with antibodies in patients with non-A  
 CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to  
 CC synthesize polypeptides, which are used to diagnose HCV-induced NANBH,  
 CC to raise antibodies for immunoassay or treatment, or to produce  
 CC vaccines. See also RAP90158, AAN90303-26, and AAN90328-36.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX

SQ Sequence 5360 BP; 1060 A; 1622 C; 1532 G; 1145 T; 1 other;

## Alignment Scores:

Pred. No.: 2.97e-72 Length: 5360  
 Score: 901.50 Matches: 175  
 Percent Similarity: 90.20% Conservative: 9  
 Best Local Similarity: 85.78% Mismatches: 11  
 Query Match: 87.35% Indels: 9  
 DB: 10 Gaps: 1

US-09-965-594-14 (1-197) x AAN90327 (1-5360)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14

DB 867 CGCAGGGCCGGAGATACTCTCGGGCCAGCCGATGGATGGTCTCCAAGGGGTGGAGG 926

QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33

DB 927 TTGCTGGGCCCCATCACGGCTAGCCGACAGCAGACAAAGGGGCCCTCTAGGGTGCATAATC 986

QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53

DB 987 ACCAGCCTAACTGGCCGGCACAACCAACAGTGGAGGTGAGTCCAGATGTGTCAACT 1046

QY 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTyrThrValThrHisGly 73

DB 1047 GCTGCCAAAGCTTCTCTGGCAACGTGCATCAATGGGGTGTCTGCACTGTCTACACGGG 1106

QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93

DB 1107 GCCGGAACGAGACCATCGCTACCCCAAGGTCCTGTCCATCCAGATGTATACCAATGTA 1166

QY 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113

DB 1167 GACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTCACACCTGCAC 1226

QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133

DB 1227 TCGGGCTCTCGGACCTTACCTGTGTACAGGACGCGCGATGTCAATCCCGTGGCGGG 1286

QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153

DB 1287 CGGGGTGATAGCAGGGGCGAGCTGTGTCGCCCGGCCCATTTCTACTTGAAGGCTCC 1346

QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173

DB 1347 TCGGGGGTCCGCTGTGTGCCCCCGGGGCGACCGCTGGGCATATTTAGGGCCCGGGT 1406

QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193

DB 1407 TGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACACC 1466

QY 194 MetArgSerPro 197

DB 1467 ATGAGGTCCCG 1478

## RESULT 13

AAN92103

ID AAN92103 standard; DNA; 6905 BP.

XX AAN92103;

AC AAN92103;

XX 25-MAR-2003 (updated)

DT 02-MAR-1990 (first entry)

XX Combined open reading frames of the hepatitis C virus (HCV) cDNAs from

[illegible]

CC It is a double-stranded nucleotide sequence of the open reading frame  
CC (ORF) (tag a) extending through clones K9-1 to 15c of hepatitis C virus  
CC (HCV) cDNA. It can be used to make oligomeric DNA hybridisation probes to  
CC detect the presence of HCV nucleic acids in samples. The polypeptide(s)  
CC it encodes could be used as immunoassay reagents and vaccines and to  
CC generate antibodies useful in diagnosis and passive immunotherapy for  
CC HCV infection/non-A, non-B hepatitis.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
XX (Updated on 25-MAR-2003 to correct PI field.)  
SQ Sequence 7310 BP; 1491 A; 2217 C; 2058 G; 1540 T; 4 other;

Alignment Scores:  
Pred. No.: 4,35e-72 Length: 7310  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
Gaps: 10

US-09-965-594-14 (1-197) x AAN92106 (1-7310)

Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 1665 CGCAGGGCGGGAGATACCTCTCGGGCCAGCCCGATGGTCTCCAGGGGTGGAGG 1724

Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 1725 TTGCTGGCGCCCATCAGCGCTAGCGCCAGACAGAGGGCCCTCTAGSGTGTATATC 1784

Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 1785 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTAGGTCAGATTGTCTCAACT 1844

Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1845 GCTGCCCAAACTTCTCGGCAACGTGATCAATGGGGTGCTGACTGTCTACACGGG 1904

Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1905 GCCGGAACGAGGACCATCGCTCACCAGGGTCTGTCTATCCAGATGTATACCAATGTA 1964

Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1965 GACCAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCATTCGACACCCCTGCAC 2024

Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 2025 TGGCGCTCTCGGACCTTACTGTGTACAGAGGACGCCGGAIGTCATTCCTCGCGCGG 2084

Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 2085 CGGGGTGATAGCAGGGGAGCGCTGTGTGCGCCCGGCCATTTCTACTTTGAAAGGCTCC 2144

Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
Db 2145 TCGGGGGTCCGCTGTGTGTGCGCGGGGACGCCCTGGGCATATTAGGGCCGCGGTG 2204

Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 2205 TGCACCGGTGGAGTGGCTAAGCGGTGACGCTTTATCCCTGTGGAGACCTTAGACAAACC 2264

Qy 194 MetArgSerPro 197  
Db 2265 ATGAGTCCCGG 2276

RESULT 15  
ID AAN90336  
XX AAN90336 standard; DNA: 7310 BP.  
AC AAN90336;  
XX 25-MAR-2003 (updated)

DT 19-JUL-2001 (updated)  
DT 01-NOV-1989 (first entry)  
DE Composite hepatitis C virus (HCV) cDNA.  
XX Hepatitis C virus; cDNA; clone 15e; clone K9-1; probe; vaccine; ds.  
XX Pan troglodytes.  
PN GB2121511-A.  
XX 26-JUL-1989.  
XX 18-NOV-1988; 88GB-0027024.  
XX 18-NOV-1987; 87US-0122714.  
PR 30-DEC-1987; 87US-0139886.  
PR 26-FEB-1988; 88US-0161072.  
PR 26-OCT-1988; 88US-0263584.  
XX (CHIR ) CHIRON CORPORATION.  
XX Houghton M, Choo QL, Kuo G;  
XX WPI; 1989-215054/30.  
DR P-PSDB; AAP90288.  
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes,  
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment  
PT of infection.  
XX Disclosure; fig 47; 235pp; English.  
XX The sequence shows a composite hepatitis C virus (HCV) cDNA, derived by  
CC aligning clones K9-1 through 15e in 5'-3' direction. The cDNA  
CC encodes antigens which react with antibodies in patients with non-A  
CC non-B hepatitis (NANBH). The cDNA can be used to design probes, or to  
CC synthesise polypeptides, which are used to diagnose HCV-induced NANBH,  
CC to raise antibodies for immunoassay or treatment, or to produce  
CC vaccines. See also AAP90288, and AAN90303-35.  
CC (N.B. This record was resubmitted to correct errors in the sequence.)  
CC (Updated on 25-MAR-2003 to correct PR field.)  
XX SQ Sequence 7310 BP; 1495 A; 2218 C; 2058 G; 1539 T; 0 other;

Alignment Scores:  
Pred. No.: 4,35e-72 Length: 7310  
Score: 901.50 Matches: 175  
Percent Similarity: 90.20% Conservative: 9  
Best Local Similarity: 85.78% Mismatches: 11  
Query Match: 87.35% Indels: 9  
Gaps: 10

US-09-965-594-14 (1-197) x AAN90336 (1-7310)

Qy 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14  
Db 1665 CGCAGGGCGGGAGATACCTCTCGGGCCAGCCCGATGGTCTCCAGGGGTGGAGG 1724

Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 1725 TTGCTGGCGCCCATCAGCGCTAGCGCCAGACAGAGGGCCCTCTAGSGTGTATATC 1784

Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
Db 1785 ACCAGCCTAACTGCGCGGACAAAACCAAGTGGAGGTAGGTCAGATTGTCTCAACT 1844

Qy 54 AlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73  
Db 1845 GCTGCCCAAACTTCTCGGCAACGTGATCAATGGGGTGCTGACTGTCTACACGGG 1904

Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVal 93  
Db 1905 GCCGGAACGAGGACCATCGCTCACCAGGGTCTGTCTATCCAGATGTATACCAATGTA 1964

Qy 94 AspLysAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 1965 GACCAAGACCTTGTGGCTGGCGCTCCGCAAGGTAGCCGCTCATTCGACACCCCTGCAC 2024

Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 2025 TGGCGCTCTCGGACCTTACTGTGTACAGAGGACGCCGGAIGTCATTCCTCGCGCGG 2084

Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
Db 2085 CGGGGTGATAGCAGGGGAGCGCTGTGTGCGCCCGGCCATTTCTACTTTGAAAGGCTCC 2144

Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
Db 2145 TCGGGGGTCCGCTGTGTGTGCGCGGGGACGCCCTGGGCATATTAGGGCCGCGGTG 2204

Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 2205 TGCACCGGTGGAGTGGCTAAGCGGTGACGCTTTATCCCTGTGGAGACCTTAGACAAACC 2264

Qy 194 MetArgSerPro 197  
Db 2265 ATGAGTCCCGG 2276



```
Db 1905 GCCGGAACGAGGACCATCGCGGTCAACCAAGGGTCCTGTCTCATCCAGATGTATACCAATGTA 1964
Qy 94 AsplysAspLeuValGlyTtpProAlaProGlnGlySerArgSerLeuThrProCysThr 113
    |||:::|||||
Db 1965 GACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTTGACACCCCTGCACT 2024
Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
    |||
Db 2025 TCGGCTCTCTCGGACCTTACCTGGTCACGAGGCACGCCGATGTCATTCCCGTGGCGCGG 2084
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
    |||
Db 2085 CGGGGTGATAGCAGGGGCGAGCTGTGTGCCCCCGGCCCATTTCTCTACTTTGAAAGGCTCC 2144
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValAlaGlyIlePheArgAlaAlaVal 173
    |||
Db 2145 TCGGGGGTCCGCTGTGTGCCCCCGGGGGCACGCCGTGGGCATATTTAGGGCCCGCGTG 2204
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
    |||
Db 2205 TGCACCCGTGCAGTGGCTAAGCGGCTTGACCTTTATCCCTGTGGAGAACCTTAGACACAACC 2264
Qy 194 MetArgSerPro 197
    |||
Db 2265 ATGAGGTCCCGG 2276
```

Search completed: August 30, 2003, 19:47:57  
Job time : 192.939 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2a model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-14  
Perfect score: 1032  
Sequence: 1 MKKGGVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+\_p2n.model -DEV=xl  
-O=/cgn2.1/USPTO\_spool/US09965594/runat\_29082003.151919.28322/app\_query.fasta.1.2872  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.rcgn.1\_12630.@runat\_29082003.151919.28322 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_nam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	106	10.3	984	10	BF304699	BF304699 601888252
c 2	103.5	10.0	1199	13	BQ892487	BQ892487 AGENCOURT
c 3	99	9.6	615	12	BJ001625	BJ001625 BJO01625
c 4	99	9.6	643	12	BJ024121	BJ024121 BJO24121
c 5	99	9.6	754	12	BJ016176	BJ016176 BJO16176
c 6	98.5	9.5	961	10	BF203316	BF203316 601865914
c 7	98.5	9.5	1403	13	BQ926101	BQ926101 AGENCOURT
c 8	98	9.5	1141	11	AK080345	AK080345 Mus muscu
c 9	97.5	9.4	779	10	BF631437	BF631437 HVSMB001
c 10	96	9.3	701	10	BF863244	BF863244 963042C02
c 11	96	9.3	846	10	BF182274	BF182274 601804028
c 12	95.5	9.3	901	10	BF307233	BF307233 601891502
c 13	95	9.2	407	9	AW785806	AW785806 117260 MA
c 14	95	9.2	931	13	BQ878887	BQ878887 AGENCOURT
c 15	94.5	9.2	641	9	AU127824	AU127824 AU127824
c 16	94.5	9.2	844	12	B1198486	B1198486 602760491
c 17	94.5	9.2	938	13	BQ894657	BQ894657 AGENCOURT
c 18	94	9.1	1283	13	BQ709745	BQ709745 AGENCOURT
c 19	93.5	9.1	701	14	CD262790	CD262790 DSM40194E
c 20	93.5	9.1	832	10	BG387051	BG387051 602454749
c 21	93.5	9.1	905	13	BU542842	BU542842 AGENCOURT
c 22	93.5	9.1	993	9	AL555424	AL555424 AL555424
c 23	93.5	9.1	1291	10	BE622016	BE622016 601440668
c 24	93	9.0	960	10	BQ955406	BQ955406 AGENCOURT
c 25	93	9.0	1146	12	BM915803	BM915803 AGENCOURT
c 26	93	9.0	1384	29	CC221189	CC221189 CH261-183
c 27	93	9.0	1637	11	AK038857	AK038857 Mus muscu
c 28	93	9.0	1702	11	AK081278	AK081278 Mus muscu
c 29	92.5	9.0	556	14	CB216999	CB216999 NISC_nq11
c 30	92.5	9.0	582	14	CB286751	CB286751 CMD45_C08
c 31	92.5	9.0	715	9	AU125614	AU125614 AU125614
c 32	92.5	9.0	846	13	BU540812	BU540812 AGENCOURT
c 33	92.5	9.0	866	13	BX451426	BX451426 BX451426
c 34	92.5	9.0	881	14	CD105862	CD105862 AGENCOURT
c 35	92.5	9.0	929	13	BQ672290	BQ672290 AGENCOURT
c 36	92.5	9.0	947	13	BU556872	BU556872 AGENCOURT
c 37	92.5	9.0	958	10	BG420860	BG420860 602452062
c 38	92.5	9.0	979	13	BQ673186	BQ673186 AGENCOURT
c 39	92.5	9.0	1008	12	BF755608	BF755608 603027112
c 40	92	8.9	871	10	BG178418	BG178418 602330206
c 41	92	8.9	898	10	BG385514	BG385514 602453808
c 42	92	8.9	963	10	BF794182	BF794182 602255566
c 43	92	8.9	1001	13	BQ928211	BQ928211 AGENCOURT
c 44	92	8.9	1640	10	BF180599	BF180599 601808704
c 45	91.5	8.9	422	14	CB763743	CB763743 AMGNNUC:S

ALIGNMENTS

RESULT 1  
BF304699/c  
LOCUS 601888252F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4122276 5',  
DEFINITION mRNA sequence.  
ACCESSION BF304699  
VERSION BF304699.1 GI:11251586  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 984)

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>;  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LCM1005 row: 9 column: 13  
 High quality sequence stop: 646.  
**FEATURES** Location/Qualifiers  
 1..984  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_image="412276"  
 /tissue\_type="rhabdomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC.17"  
 /note="Organ: muscle; Vector: pOTB7; Site:1: EcoRI;  
 Site:2: XhoI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
**BASE COUNT** 133 a 329 c 351 g 171 t  
**ORIGIN**

Alignment Scores:  
 Pred. No.: 4.67 Length: 984  
 Score: 106.00 Matches: 33  
 Percent Similarity: 45.24% Conservative: 5  
 Best Local Similarity: 39.29% Mismatches: 24  
 Query Match: 10.27% Indels: 22  
 DB: 10 Gaps: 5

US-09-965-594-14 (1-197) x BF304699 (1-984)

Qy 100 TrpProAlaProGlnGlySerArgSerLeuThr----ProCysThrCysGlySerSerAsp 118  
 Db 646 TGGCCCCAGTCCAGCGCATCCGGTGGGAGAGAGACCGGTACCTGC-----599  
 Qy 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138  
 Db 598 -----ACCAGGACGACGCAACATACATCAAGAGACGTGCT---TCCGCG 554  
 Qy 139 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySer 153  
 Db 553 GGGCGCCTCTGTGGGAGAACCTCGATGGTGTCAAGCTCGCGCTGCTACTGGAAGT 494  
 Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
 Db 493 CGGCACGCTCGGTGAGTGCAGC-----TTCCAGCGCCGGGG 455  
 Qy 174 CysThrArgGly 177  
 Db 454 TGGCGCGGAGGA 443

RESULT 2  
 BQ892487  
 LOCUS BQ892487 1199 bp mRNA linear EST 16-AUG-2002  
 DEFINITION AGNCOURT\_8417538 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
 IMAGE:6192708 5', mRNA sequence.  
 ACCESSION BQ892487  
 VERSION BQ892487.1 GI:22284501  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 1199)  
**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>;  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM13595 row: c column: 13  
 High quality sequence start: 57  
 High quality sequence stop: 394.  
**FEATURES** Location/Qualifiers  
 1..1199  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_image="6192708"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski\_sympathetic\_trunk"  
 /note="Vector: PCMV-SPORT6 (Life Technologies); Site:1:  
 NotI; Site:2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TGACCCACGCGTCG-3' and  
 5'-GACTAGTTCGTAGTCGAGGCGCGCT(15)-3'. Size selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine); available through Life  
 Technologies."  
**BASE COUNT** 255 a 362 c 343 g 211 t 28 others  
**ORIGIN**

Alignment Scores:  
 Pred. No.: 10.4 Length: 1199  
 Score: 103.50 Matches: 41  
 Percent Similarity: 37.42% Conservative: 17  
 Best Local Similarity: 26.45% Mismatches: 53  
 Query Match: 10.03% Indels: 44  
 DB: 13 Gaps: 6

US-09-965-594-14 (1-197) x BQ892487 (1-1199)

Qy 68 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIle 87  
 Db 484 TGGGATCCATTTTAAATAAGGGTGCTCTGTATATCGGCCGCCACGGCGCGTGATA 543  
 Qy 88 GlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGlnGlySerArg 107  
 Db 544 CTTCCATTATACCACATGTGACAGTCACTTT-----573  
 Qy 108 SerLeuThrProCysThr-----CysGlySerSerAsp 118  
 Db 574 ---TGTGCTGCCTGCACACACCCCATCCGATGTTGGCGCTATTGTGGAACGGCGAG 630  
 Qy 119 LeuTyr-LeuValThr-----ArgHisAlaAspValIleProValArg----- 132  
 Db 631 CGGTTTCATTGGCCACTCCCTCTCTATAAAACACGCCAACGTCGTTCATGGGCGGGCT 690  
 Qy 133 -----ArgArgGlyAspSerArgGlySerLeuLeu-- 142  
 Db 691 GGGTGTGTTGGCAGCGCAAGCGGGGTGGGGCATGTTAGGACTCGGGGGCGGATCTCTG 750  
 Qy 143 -----SerProArgProIleSerTyrLeuLys-----GlySerSerG1 155

```

Db      751 AAAACCCACCCTCGGCCCCACCGATGCGCTAAGCTCCCTTTACAGCCACCGCCGCGG 810
      155 yGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysTh 175
      811 CCCCCTTAACTATCTCTACCTCGCGCGCGCGGGGGGAGACAGTGGCGCATACGGGC 870
      175 rArgGlyValAlaLysAlaValAspPheIleProValGluSer 189
      871 TCAGGCGCTTTTAAAGCCCGCGGCTTCGCGCGCGGCGGAAGCA 913

RESULT 3
BJ001625/c
LOCUS      BJ001625 615 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
            mRNA sequence.
ACCESSION  BJ001625
VERSION     BJ001625
KEYWORDS   BJ001625
SOURCE     BJ001625.1 GI:17364516
ORGANISM   Oryzias latipes (Japanese medaka)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 615)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1. .615
               /organism="Oryzias latipes"
               /mol_type="mRNA"
               /strain="Hd-rR"
               /db_xref="taxon:8090"
               /clone="MF01SSA025C02"
               /sex="mixture of female and male"
               /tissue_type="whole embryo"
               /dev_stage="segmentation stage 20 - 25"
               /clone_lib="MF01SSA cDNA"
BASE COUNT 140 a 156 c 165 g 144 t
ORIGIN

Alignment Scores:
Pred. No.: 12 Length: 615
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.59% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-14 (1-197) x BJ001625 (1-615)

Qy      41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
      511 AAAAATGACGTAGAACCAAAAGACACAGATCCACACACATCTCTGCTTCACGGCT 452
      61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80
      451 -----TGTGGAGAACCTATCACAGTTCCTGCTTTAGACGACGCA 410
      81 SerProLys-----GlyProValIleGlnMetTyThrAsnValAspLys 95
      409 GCTCTGCGCGCGGAGGAGCTCTCGGCCAGTTGTG----- 374
      96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro----- 111

```

```

Db      373 -----ACTCGTGAGAGCAAGAGCGTCACCCCGAGCTGAGG 335
      112 -----CysThrCysGlySerSerAspLeuTyIleuValThrArg----- 124
      334 CTGACAGGATCGGATGTGGCTCTGCT-----TTGGTTCTCTGCTCTCTCGGATCA 284
      125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
      283 TCTTCTCACCTGACCTTCCACATCCAGGTGTGCGCAGCGCTGTCTGACGGGTGATGG 224
      138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyLeuLysGlySerSer 154
      223 AGAGCCCGGACAGCAGCAGTCGGGGTGAATCTCTGCAGGACGCTCTTCACGGCGGATCA 164
      155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
      163 GGAGGACCGACTCGCTGCAGAGCCCTCTGCTGCA 131

RESULT 4
BJ024121
LOCUS      BJ024121 643 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ024121 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA143D12 3',
            mRNA sequence.
ACCESSION  BJ024121
VERSION     BJ024121
KEYWORDS   BJ024121
SOURCE     BJ024121.1 GI:17377389
ORGANISM   Oryzias latipes (Japanese medaka)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 643)
AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE     Medaka EST Project in Takeda's lab
JOURNAL   Unpublished
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
            Location/Qualifiers
            1. .643
               /organism="Oryzias latipes"
               /mol_type="mRNA"
               /strain="Hd-rR"
               /db_xref="taxon:8090"
               /clone="MF01SSA143D12"
               /sex="mixture of female and male"
               /tissue_type="whole embryo"
               /dev_stage="segmentation stage 20 - 25"
               /clone_lib="MF01SSA cDNA"
BASE COUNT 171 a 148 c 148 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 12.7 Length: 643
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 9.59% Indels: 50
DB: 12 Gaps: 7

US-09-965-594-14 (1-197) x BJ024121 (1-643)

Qy      41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
      242 AAAAATGACGTAGAACCAAAAGACACAGATCCACACACATGTTCTGCTTACGGCT 301
      61 ThrCysIleAsnGlyValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAla 80

```

```

Db 302 -----TCTTGAGAACCTATCACAGTTCTCTGTAGACGACGGCA 343
Qy 81 SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 95
Db 344 GCTCTGGCGGGGAGGAGCTCTGGCCAGTTGTG-----
Qy 96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro-----111
Db 380 -----ACTCGTGAGGAGGAGGAGGCTCACCCGAGGCTGTAGG 418
Qy 112 -----CysThrCysGlySerSerAspLeuTyrLeuValThrArg-----124
Db 419 CTGACGGATCGGATGTGGCTCTGCT-----TTGGTTCTCTCTCTCTGGATCA 469
Qy 125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
Db 470 TCTTCTCCTCCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGTGACGGGTGATGG 529
Qy 138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 154
Db 530 AGAGCGGACAGCAGCAGTGGGGGTGAATCTCTGCAGGACGCTCTTCACGGCGGATCA 589
Qy 155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
Db 590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 5
BJ016176 754 bp mRNA linear EST 05-DEC-2001
LOCUS BJ016176
DEFINITION BJ016176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',
mRNA sequence.
ACCESSION BJ016176
VERSION BJ016176.1 GI:17376695
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 754)
AUTHORS Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..754
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA025C02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"
BASE COUNT 194 a 181 c 181 g 198 t
ORIGIN

Alignment Scores:
Pred. No.: 15.6 Length: 754
Score: 99.00 Matches: 42
Percent Similarity: 33.77% Conservative: 9
Best Local Similarity: 27.81% Mismatches: 50
Query Match: 95.9% Indels: 50
DB: 12 Gaps: 7

```

```

US-09-965-594-14 (1-197) x BJ016176 (1-754)
Qy 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAla 60
Db 242 AAAAATGAGCTGTAGAACCAAAACACACAGATCCACACACATGTTCTGTCTACGGGCT 301
Qy 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 302 -----TGTTGAGAACCTATCACAGTTCTCTGTAGACGACGGCA 343
Qy 81 SerProLys-----GlyProValIleGlnMetTyrThrAsnValAspLys 95
Db 344 GCTCTGGCGGGGAGGAGCTCTGGCCAGTTGTG-----
Qy 96 AspLeuValGlyTrpProAlaProGlnGlySerArgSerLeuThrPro-----111
Db 380 -----ACTCGTGAGGAGGAGGAGGCTCACCCGAGGCTGTAGG 418
Qy 112 -----CysThrCysGlySerSerAspLeuTyrLeuValThrArg-----124
Db 419 CTGACGGATCGGATGTGGCTCTGCT-----TTGGTTCTCTCTCTGGATCA 469
Qy 125 -----HisAlaAspValIleProValArgArgGlyAspSer 137
Db 470 TCTTCTCCTCCTGACCTTCCACATCCAGGTGTCGCCAGCGCTGTGACGGGTGATGG 529
Qy 138 ArgGlySerLeuLeuSerProArg-----ProIleSerTyrLeuLysGlySerSer 154
Db 530 AGAGCGGACAGCAGCAGTGGGGGTGAATCTCTGCAGGACGCTTCACGGCGGATCA 589
Qy 155 GlyGlyProLeuLeuCysProAlaGlyHisAla 165
Db 590 GGAGGACCGACTCGCTGCAGAGCCTCTGCTGCA 622

RESULT 6
BF203316 961 bp mRNA linear EST 06-NOV-2000
LOCUS BF203316
DEFINITION BF203316 NTH_MGC_17 Homo sapiens cDNA clone IMAGE:4098578 5',
mRNA sequence.
ACCESSION BF203316
VERSION BF203316.1 GI:11096902
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 961)
AUTHORS NTH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM965 row: 1 column: 03
High quality sequence stop: 637.
FEATURES
Location/Qualifiers
1..961
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4098578"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/notes="Organ: muscle; Vector: pOTB7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the

```

following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). "

230 a	297 c	300 q	134 t
-------	-------	-------	-------

BASE COUNT

Alignment Scores:		
Pred. No.:	23.8	Length:
Score:	98.50	Matches:
Percent Similarity:	52.44%	Conservative:
Best Local Similarity:	39.02%	Mismatches:
Query Match:	9.54%	Indels:
DB:	10	Gaps:
		961

US-09-965-594-14 (1-197) x BF203316 (1-961)

Qy		101	ProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSer-AspLeuTy	120
Db		642	CGGGTCTCTCTCCCTCCGGCA-----TGAGAAATGTGGAGCTCAGACATGCT	592
Qy		120	rLeuValThrArgHISAlaAspValIleProValArgArgGlyAspSerArgGlySe	140
Db		591	GACTCTGAGTCGACGAGCGGCACAGATCGGATCGGAACGCCGGGATGCTCATGGGG	532
Qy		140	rLeuLeuser-ProArgProIleSerTyrLeuIysGlySerSofrGlyIleProLeuLeuC	160
Db		531	CATGTCGTCTCTCTGTCCTCCCGGGGAGTAGGCCAAGGGTCTGGAGCAGGAGGS	478
Qy		160	ysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaIal	180
Db		477	GTCTCGCACCTGCTGCTTAGTGCTCTGCCCGTCAGGCTTTTGTGATGGTGGTGACGCTG	418
Qy		180	ys 180	
		:	:	
Db		417	AA 416	

## RESULT 7

BQ26101/C	BQ926101	1403 bp	linear	EST 20-AUG-2002
LOCUS	AGENCOURT_8752655	NIH_MGC_130	Mus musculus	cDNA clone IMAGE:6335718
DEFINITION	5', mRNA sequence.			

ACCESSION	BQ926101
VERSION	BQ926101.1
KEYWORDS	GI:22341132 EST.
SOURCE	Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1403)  
REFERENCE  
NIH-MGC <http://mgc.nci.nih.gov/>  
AUTHORS  
TITLE  
Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished  
COMMENT  
Contact: Robert Strausberg, Ph.D.

**COMMENT**  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-re@mail.nih.gov](mailto:cgabbs-re@mail.nih.gov)  
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.

CDNA Library Preparation: ResGen, Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: WGC clone distribution information can  
be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

```

recp://image:1111.90v
Plate: LLAM13798 row: j column: 07
High quality sequence stop: 101.

```

FEATURES	Location/Qualifiers
source	1. .1403

```

1. 1403
source
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="DH10B (phage-resistant)"

```

```

/clone_lib="NIH_MGC_130"
/notes="Organ: otcysts; Vector: pCMV-SPORT6.1 ccdb;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dr. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH MGC library."

```

BASE COUNT	297 a	521 c	237 g	345 t	3 others
ORIGIN					

Alignment Scores:	38.5	Length:	1403
Pred. No.:	98.50	Matches:	49
Score:	36.75%	Conservative:	12
Percent Similarity:	29.52%	Mismatches:	62
Best Local Similarity:	29.52%	Indels:	43
Query Match:	5.4%	Gaps:	9
DB:	13		

US-09-965-594-14 (1-197) x BQ926101 (1-1403)

Qy	11 GlyArgIleAsnLeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlucGly 30
Db	1378 GGGGTGTGTCANCGGTACAGGACAGGTGCC---GCACACTCGAGCTCGCGCCAGAGACT 1322

QY 31 CysGlnGluThr-----SerGlnThrGlyArgAspLysAsnGln-----val 44

QY 45 GluglyGluValGlnIleValSerThrAlaAAlaGlnThrPheLeuAlaThrCysIleAsn 64  
 |||||...  
 QY 1321 TGTCGGGGCGCGCTTGGCGCATACCCGGGTCGATCGAGGTGAGCGCGCTGTATACA 1362  
 DB 1321 TGTCGGGGCGCGCTTGGCGCATACCCGGGTCGATCGAGGTGAGCGCGCTGTATACA 1362

Db	1261	GAGGGGAAA	-----CAG	1350
Qy	65	GlyValCysTrpThrVal	HisGlyAlaGlyThrArg	ThrLeuAlaSerProLys--G
				84

DB  
1249 GGGTA---TGGTTATCAGCGGTGGGCGAGTACT-----TCCCTTAAGCG 1205

Db 1204 GCGSCGTGCGAGTATATATACCGGCGAGTGCAGGACGCGGCGTGGACCTTGACC 1145

Qy	104	lnGlySerArgSerLeuThrProCysGlySerSerAspLeuTyrLeuValThrA	124
Db	1144	AA----GAGAGGCACTGAGCGCCCTCCCTCTGGCGCTGTGATAATACAAATGTGCAG	1088

QY	124	rgHisAlaaspvalilleProValArgArgGlyAsp-----	136
----	-----	---	-----

DB	QY	Sequence	Score
1087	1087	GGCACGGTGATGTGGTTACTACGGCGGCACCGGCTCCAAACGGGGCTCTCTAACAGACGC	1028
137	137	-----SerArgGlySer-----LeuLeuSerProGlySerTyrLeuLysG	152

Db 1027 CCGCCTCCGCGGCAACAGGGTAATAATCATCGCGGGCGGGATTTCGCATTCGCGCGG 968  
 QY 152 lyserSerGlyGly 156

Db 967 GAGAAGGCGGGT 954

## RESULT 8

AK080545  
LOCUS

DEFINITION Mus musculus 7 days enriched library. cl

finger protein (frag  
enriched library, c  
finger protein (frag  
AK080545

ACCESSION	AKU80545
VERSION	AK080545.1
KEYWORDS	UTM, CAP, trans

KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

ORGANISM *Mus musculus*  
Eukaryota; Metazoa;

REFERENCE 1

**AUTHORS** Carninci, P. and Haya  
**TITLE** High-efficiency full

JOURNAL Meth. Enzymol. 303,  
MEDLINE 99279253

AK080545	1141 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS				
DEFINITION				
Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone: A730082L10 product: weakly similar to zinc finger protein (fragment) [Mus musculus], full insert sequence.				

ACCESSION AK080545  
VERSION AK080545.1 GI:26348600  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)

ORGANISM	MUS MUSCULUS
REFERENCE 1	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

<b>AUTHORS</b>	Carninci, P. and Hayashizaki, Y.
<b>TITLE</b>	High-efficiency full-length cDNA cloning
<b>JOURNAL</b>	Meth. Enzymol. 303, 19-44 (1999)
<b>MEDLINE</b>	99279253

PUBMED  
REFERENCE

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

20499374

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

20530913

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmshing, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

21085660

11217851

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563-573 (2002)

6 (bases 1 to 1141)

6

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanazaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, M., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

PUBMED  
REFERENCE

10349636

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

20499374

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

20530913

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmshing, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y.

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/.

## FEATURES

Location/Qualifiers

source

1..1141

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM,DB:A730082L10"

/db\_xref="taxon:10090"

/clone="A730082L10"

/tissue\_type="cerebellum"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="7 days neonate"

&lt;1..587

/note="unnamed protein product: putative

weakly similar to zinc finger protein (fragment) [Mus

musculus] (PIR|I48722, evidence: FASTY, 50.7%ID,

57.6%length, match=601)"

/codon\_start=3

/protein\_id="BAC37940.1"

/db\_xref="GI:26348601"

/translation="DSCLPASRSLITPRGDFELKELSAARAVGPGSPVAFQVS

TRYGAQAQQQRRVGRACRSEGLSKSRPRQRHVPVPGHYIGSGRRIPPPAGE

AQAAGRAPQVPHPPGHPTGVPPQAGLLPALAARQVPGVPRGREGPRAPRHS

PKPVPTALGFSFGGSGPAPPLAPANGRSVGLAL"

1118..1123

polyA\_signal

polyA\_site

1141

/note="putative"

BASE COUNT 244 a 316 c 353 g 228 t

ORIGIN

Alignment Scores:

Pred. No.: 33 Length: 1141

Score: 98.00 Matches: 46

Percent Similarity: 35.85% Conservative: 11

Best Local Similarity: 28.93% Mismatches: 52

Query Match: 9.50% Indels: 50

DB: 11 Gaps: 9

US-09-965-594-14 (1-197) x AK080545 (1-1141)

Qy 46 GlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGly 65

Db 303 GGAGAGCGCAGCGAGCTGGACGCGCGCGAG-----CGCCACATGGAACA 371

Qy 66 ValCysTrpThrValTyHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyPro 85

Db 336 -----CAAGTCCCGCATCCCGCTGGG-----CCTCCTCAAGGA 389

Qy 86 ValIleGlnMetTyThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGly 105

Db 372 GTGCTG-----CCTCCTCAAGGA 389

Qy 106 SerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyLeuValThrArgHis 125

Db 390 CGCGCTGGCTTCTCCCTGCA-----CTCGCAGCTCGCAA 425

Qy 126 AlaAspValIleProValArg---ArgArgGlyAspSerArgGlySerLeuLeuSerPro 144

Db 426 GTTCTCTGTGACCGGTTAGGGCGCGGAGGAGCAAGAGGAGGCGCAGACACAGCCCC 485

Qy 145 ArgProIle-----SerTyLeuLysGlySerGlyGlyProLeu 158

Db 486 AAGCGGTTCTTACAGCCTTGGGGTTTCGTTGGGAGGGTGGGCTGCTCTCCCTC 545

Qy 159 LeuCysProAla---GlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGly 177





Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

## FEATURES

source

1. 701  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
normalized, Lambda Zap II"  
/note="vector: pBluescript II SK-; Site.1: EcoRI; Site.2:  
XhoI; this library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoRI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t  
ORIGIN

## Alignment Scores:

Pred. No.: 27.6 Length: 701  
Score: 96.00 Matches: 32  
Percent Similarity: 40.71% Conservative: 14  
Best Local Similarity: 28.32% Mismatches: 45  
Query Match: 9.30% Indels: 22  
DB: 10 Gaps: 4

US-09-965-594-14 (1-197) x BF863244 (1-701)

Qy 71 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 86  
:::|||||  
Db 171 CACCACCATACCTTCCTCGTCTGCACACCAAAATTTATGCCCATACGGCCACTA 230  
Qy 87 IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrProAlaProGlnGlySer 106  
:::|||||  
Db 231 ACNAGTTACATACAGG-----AAGACACAGCGCGCTTGGCCACCCCTTGGAGCG 284  
Qy 107 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 126  
|||||  
Db 285 AGAAGCGGACGCGTGGTCTCTGGGTCTATCGCATCTCCCGCTATCAG 344  
Qy 127 AspValIle-----ProValArgArgArgGlyAspSerArg----- 138  
:::|||||  
Db 345 GAGATCATTTGCATGTGCGCTTTAGTCACCCCAAGAGAGCGCTGGGAGTGGGCATTATAA 404  
Qy 139 -----GlySerLeuLeuSerProArgProIleSerTyrLeu 150  
Db 405 GAAGGGGAGCGGAATTCGTTTGGGAAAGAGTACGCGCCCAAGGCTGTGACCAAGTCTGA 464  
Qy 151 LysGlySerGlyGlyProLeuLeuCysProAlaGly 163  
|||  
Db 465 CTCGAAGGACGCAATGGGAGCGCTTTCGGGTCTGCGGCT 503

## RESULT 11

BF182274/c

LOCUS 601804028F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',  
846 bp mRNA linear EST 31-OCT-2000  
DEFINITION mRNA sequence.

ACCESSION BF182274

VERSION BF182274.1

KEYWORDS GI:11060416

SOURCE Est.

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 846)

## AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@email.nih.gov](mailto:cgabbs@email.nih.gov)

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: L1AM9308 row: g column: 07  
High quality sequence stop: 996.  
Location/Qualifiers

## FEATURES

source

1. 846  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4035102"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;  
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

BASE COUNT 176 a 218 c 241 g 210 t  
ORIGIN

## Alignment Scores:

Pred. No.: 35.1 Length: 846  
Score: 96.00 Matches: 46  
Percent Similarity: 47.20% Conservative: 13  
Best Local Similarity: 36.80% Mismatches: 36  
Query Match: 9.30% Indels: 31  
DB: 10 Gaps: 9

US-09-965-594-14 (1-197) x BF182274 (1-846)

Qy 73 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAs 92  
|||||  
Db 757 GGTTCCTCTACCAGAACACGCTGGATGAAGAAAGGACCA-----CATCCTTC 710  
Qy 92 nValAspLysAspLeuValGlyTyr-----ProAlaProGl 104  
:::|||||  
Db 709 GGTTCTTCAGTCCCACTGGCTGGGAGAAAGTGAACACAGAGAAAGAGGACGGTCCCTCA 650  
Qy 104 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuVal 122  
|||  
Db 649 GTCCCCACGGTTAGTACTAGTACAGAGTCTTCTGCTGGA----- 610  
Qy 122 lThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 142  
|||||  
Db 609 -ACTAGACACACCT--GTAATCCAGGAGGAAACGCTGGAGGAACAGAGGACTCC---CT 556  
Qy 142 uSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeu---LeuCysPr 161  
:::|||||  
Db 555 GACCCCACTCC---TCCCTCTAGCGGCACCTCTCTGGCCCACTCCCTCTGTCTCC 499  
Qy 161 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 176  
:::|||||  
Db 498 TAGTGGGCACCTCTCCCGACGACACAGACTGTACTCCCTTTGGCCCTCTGCACCT 439  
Qy 176 gGlyValAlaLys 180  
|||  
Db 438 TGGGATGACTGAG 426

## RESULT 12

BF307233

```

LOCUS       BF307233               901 bp      mRNA      linear      EST 21-NOV-2000
DEFINITION   601891502F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137145 5',
            mRNA sequence.
ACCESSION    BF307233
VERSION      BF307233.1   GI:11254342
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 901)
AUTHORS      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L1C1044 row: C column: 02
            High quality sequence start: 6
            High quality sequence stop: 684.
            Location/Qualifiers
                1..901
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4137145"
                /tissue_type="rhodomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_17"
                /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
            Site_2: XhoI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size-selected >500bp
            for average insert size 1.8kb. Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 144 a 267 c 329 g 161 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:
66	42.5	95.50	37
Percent Similarity:	38.46%	Conservative:	8
Best Local Similarity:	31.62%	Mismatches:	28
Query Match:	9.25%	Indels:	45
DB:	10	Gaps:	5

US-09-965-594-14 (1-197) x BF307233 (1-901)

Qy 66 ValCysTrpThr-ValTyrHis-----GlyAlaGlyThrArgThrIleAl 80  
 ::::|||||::: ||| |||  
 620 ATGTGTTGGAGCGTCCCGCACGGCATCTGAGCGGGTCTCGGCACACACACGCTGG 679  
 ::::|||||::: ||| |||

Qy 80 aSerProLysGlyProValIleGlnMetTyrThrAsnValAspLysAspLeuValGlyTr 100  
 :||| |||  
 680 TGGCGCAGGAGCGTGGAGTGTGTGCA-----GTGTCAGGATG 715  
 :||| |||

Qy 100 pProAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120  
 :||||| |||  
 716 GCCCGCCCATCCGG-----731

Qy 120 rLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySe 140  
 :||| |||  
 732 -----GTGAGCCCTCGTCTCAGGCGGCTTGGCGGGGTTTC 766  
 :||| |||

Qy 140 rLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCy 160  
 :||||| |||  
 :||| |||

```

Db 767 CCTTTTGGTCTCTGA-----CGGGCATCTCTCCAGGGCGCGCTGGACTG 810
Qy 160 sprAlaGlyHisAlaValGly-----IlePheArgAlaAlaValCys 174
  :||||| ||| |||
  811 TCCGGCCGGTTCGCCCGGGCGGCCACACAGGGTCCGGCGGCGCTGTGC 859
  :||||| ||| |||

RESULT 13
AW785806
LOCUS       117260 MARC 1P1G Sus scrofa cDNA 5', mRNA linear EST 09-JUL-2000
DEFINITION   AW785806
ACCESSION    AW785806
VERSION      AW785806.1   GI:7842582
KEYWORDS     EST.
SOURCE       Sus scrofa (pig)
ORGANISM     Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE    1 (bases 1 to 407)
AUTHORS      Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Perlea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
            porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
            Mamm. Genome 13 (8), 475-478 (2002)
JOURNAL      Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE      22213789
PUBMED       12226715
COMMENT      Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithdemail.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAI
            BACKWARD: GTTTCCTCCAGTCACGAG
            Plate: 37 row: D column: 16
            Seq primer: ATTTAGGTGACACTATAG.
            Location/Qualifiers
                1..407
                /organism="Sus scrofa"
                /mol_type="mRNA"
                /db_xref="taxon:9823"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /clone_lib="MARC 1P1G"
                /note="Vector: pCMV SPOR6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from day 11, 13, 15, 20,
            and 30 embryos."
```

BASE COUNT 55 a 131 c 145 g 76 t

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:
43	17.2	95.00	41
Percent Similarity:	31.21%	Conservative:	3
Best Local Similarity:	29.08%	Mismatches:	52
Query Match:	9.21%	Indels:	45
DB:	9	Gaps:	7

US-09-965-594-14 (1-197) x AW785806 (1-407)

Qy 43 GlnValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCys 62  
 :||| |||  
 61 CAACCCCGGCGAGCTCTTCGGCCGATCCCTATCGCTTCTCGGTGATGCGAGATGC 120  
 :||| |||

Qy 63 IleAsnGlyValCysTrpThrVal-----TyrHis-Gl 73  
 :||| |||  
 121 GTTTCGACGGCTGTCTGGGCTTCCGGGGGCTTCGTGCACCGGCGCTTCTCGGTGATGG 180  
 :||| |||

Qy 73 yAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMetTyrThrAsnVa 93

```

Db      181 AGGACGGACTAACCGGGTCTGGGCT-----GGGCT----- 214
QY      93 LasplysAspLeuValGlyTrpProAlaProGlnGlySerArg-----SerLeuThrPr 111
Db      215 -----GGGCTGGCTGGCTCAGCGGAGGAGAGACTACTTAAGCTCACACC 258
QY      111 cYThrCysGlySerSerAspLeuValThrArgHis----- 125
Db      259 T-----GACCGAGGCGCCGACCGCGCTGGCGCACCTGTACGGC 300
QY      126 ----AlaaspValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerPr 144
Db      301 GCGAGCTACCGCTCGAGCAGCTGCACCGCTGGAGATCAGCGGGTGCACTCGCGGAC 360
QY      144 cArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCysProAlaGlyHI 164
Db      361 ACGGCT-----GGAGGTGCTGGGACTTGTGCGTGTCCCACTGTACA 402
QY      164 s 164
Db      403 C 403

RESULT 14
LOCUS   BQ878887
DEFINITION AGENCOURT_8119707 Lupski_dorsal_root_ganglion Homo sapiens cDNA
clone IMAGE:617774 5', mRNA sequence.
ACCESSION BQ878887
VERSION   BQ878887.1 GI:22270895
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 931)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bhs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13556 row: e column: 07
High quality sequence start: 18
High quality sequence stop: 705.
Location/Qualifiers
1..931
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:617774"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/clone_lib="Lupski_dorsal_root_ganglion"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
Not1; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adapters:
5'-TCGACCTTCAGTCCGCG-3' and
5'-GACTAGTTCAGTCCGCGCGCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 184 a 308 c 253 g 183 t 3 others

```

```

ORIGIN
Alignment Scores:
Pred. No.: 49.4 Length: 931
Score: 95.00 Matches: 46
Percent Similarity: 40.13% Conservative: 15
Best Local Similarity: 30.26% Mismatches: 52
Query Match: 9.21% Indels: 40
DB: 13 Gaps: 7

US-09-965-594-14 (1-197) x BQ878887 (1-931)
QY 10 ValGlyArgIleAsnLeuSerGlyAspThrAlaGlnThrArgGlyGluGlu 29
Db 235 ATTGGGNAATCAATCTTAGTGGAGGACCACTACACCCGTCCTCAGGAGT----- 288
QY 30 GlyCysGlnGlnThrSerGlnThrGlyArgAspLysAsnGlnValGlyGluGluValGln 49
Db 289 -----CGAGTCACCATTTTCAGCAGACACGTCACCAAGAACCGGTCTCCCTGAAGCTGAGC 342
QY 50 lIeValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCys----- 67
Db 343 TCTGTG----ACCGCGCGGACACGCTCTGTATTACTGTGCGAGAGGTGCTTTCGTTG 399
QY 68 ---TrpThrValTyrHisGlyAlaGlyThrArg-----ThrIleAlaSerProLys 83
Db 400 TACTACTTTGACTACTGGGCGCAGGAACTGTCACCGTCTCTCAGCTCCACCAAG 459
QY 84 GlyProValIleGlnMetTyrThrAsnValAspLysLeuValGlyTyrProAlaPro 103
Db 460 GCGCCATCGGTC-----TTCCCTC----- 477
QY 104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr----- 120
Db 478 -----CTGGCGCGCTGCTCCAGGAGCACCTCCGAGAG-CACAGCGGCGCT 521
QY 121 -----LeuValThrArgHisAlaaspValIleProValArgArgArg 134
Db 522 GGGCTGCTGTGTCAGGAGACTACTTCCCGAAACCGGTGTCGTGGAACATCAGCGC 581
QY 135 GlyAspSerArgGlySerLeuLeuSerProArgPro 146
Db 582 TCTGACGAGCGCGTGCACACCTTCCCGAGCTGCTCT 617

RESULT 15
LOCUS   AU127824
DEFINITION AU127824 NT2RP2 Homo sapiens cDNA clone NT2RP2002160 5', mRNA
sequence.
ACCESSION AU127824
VERSION   AU127824.1 GI:10988178
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 641)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```

```

FEATURES
  source      Location/Qualifiers
1..641
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2002160"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP2"
/note="Vector: pME18SFL3: mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT      103 a   256 c   183 g   96 t       3 others
ORIGIN

Alignment Scores:
Pred. No.:      34.3      Length:      641
Score:          94.50      Matches:    47
Percent Similarity: 39.10%      Conservative: 14
Best Local Similarity: 30.13%      Mismatches: 56
Query Match:      9.16%      Indels:    39
DB:               9      Gaps:      9

US-09-965-594-14 (1-197) x AUI27824 (1-641)

QY   67  CysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
Db   154  TGCACGACGACACCCCTACCGCGGGGAGCAGCCACCCACCCACCGCGGGCCCTGGC 213
QY   87  IleGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpProAlaProGlnGlySer 106
Db   214  GGGCAATG-----ACATCCTGGCCAGCCCCCTCGCCTGC 249
QY   107  ArgSerLeuThrPro-----CysThrCysGlySer 116
Db   250  CCAGCCCCAGCCCTACCCCGGAGCCCGCCACACAGCTCCTACGTGCACCTGCGGCCGG 309
QY   117  SerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp 136
Db   310  CACGAC-----CCACAAGCCACCCG-----CCACAGCCACCGGACTTC 351
QY   137  SerArgGlySer-----LeuLeuSerProArgProIleSerTyrLeuLysGlySerSer 154
Db   352  AGCCGGTGCTCCACCTGGTTGCGCTCAACAGCCCCCTGTGTCAGCGGCATGCGGGCATCC 411
QY   155  GlyGlyPro-Leu---LeuCysProAlaGly-HisAlaValGlyIle-----PheA 170
Db   412  GCGGGCGGACTTCCAGTGCTTCCAGCAGCGCGGGCCGTGGGGCTGGCGGGCACCTTC 471
QY   170  rgAlaAlaValCysThrArg-----GlyValAlaLysAlaValAsp---- 183
Db   472  GCACCTTCCTGCTCTCGCCTGCAGGACCTGTACGGCATGTCGGCGTGGCGGCGCG 531
QY   184  --PheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db   532  CAACCGTGGCCATGCTCAACCTCAAGGACGAACTGCTGTTTCCA 575

```

Search completed: August 31, 2003, 04:27:28  
Job time : 1914.31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 : Search time 44.6227 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKKGSWIVGRINLSGDTA.....YAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032	100.0	197	21	AA15222 Hepatitis C virus
2	1015	98.4	197	21	AA15221 Hepatitis C virus
3	1002	97.1	197	21	AA15223 Hepatitis C virus
4	990	95.9	197	21	AA15224 Hepatitis C virus
5	981	95.1	195	21	AA15220 Hepatitis C virus
6	980	95.0	197	21	AA15225 Hepatitis C virus
7	976	94.6	197	21	AA15226 Hepatitis C virus
8	942	91.3	195	21	AA15212 Hepatitis C virus
9	911.5	88.3	665	20	AA124943 HCV NS4A-NS3 compl

ID	908.5	88.0	665	20	AA124947	HCV NS4A-NS3 compl
11	907.5	87.9	665	20	AA124941	HCV NS4A-NS3 compl
12	907.5	87.9	665	20	AA124942	HCV NS4A-NS3 compl
13	904.5	87.6	665	20	AA17880	HCV NS4A-NS3 compl
14	904.5	87.6	665	20	AA124945	HCV NS4A-NS3 compl
15	904.5	87.6	665	20	AA124946	HCV NS4A-NS3 compl
16	903.5	87.5	665	20	AA124940	HCV NS4A-NS3 compl
17	903.5	87.5	671	20	AA124948	HCV NS4A-NS3 compl
18	901.5	87.4	216	20	AA17884	HCV NS4A-NS3 compl
19	900.5	87.3	216	20	AA17879	HCV NS4A-NS3 compl
20	900.5	87.3	216	20	AA17878	HCV NS4A-NS3 compl
21	900.5	87.3	665	20	AA124944	HCV NS4A-NS3 compl
22	900.5	87.3	671	20	AA124949	HCV NS4A-NS3 compl
23	900	87.2	215	20	AA17890	HCV NS4A-NS3 compl
24	897.5	87.0	216	20	AA17882	HCV NS4A-NS3 compl
25	897.5	87.0	216	20	AA17883	HCV NS4A-NS3 compl
26	897.5	87.0	216	20	AA17886	HCV NS4A-NS3 compl
27	896.5	86.9	216	20	AA17877	HCV NS4A-NS3 compl
28	894	86.6	215	20	AA17887	HCV NS4A-NS3 compl
29	893.5	86.6	216	20	AA17881	HCV NS4A-NS3 compl
30	893.5	86.6	216	20	AA17885	HCV NS4A-NS3 compl
31	889	86.1	213	20	AA17888	HCV NS4A-NS3 compl
32	889	86.1	631	20	AA193482	HCV NS3 protein.
33	888.5	86.1	3011	19	AA177397	Hepatitis C virus
34	888.5	86.1	3011	24	ABP71460	Amino acid sequenc
35	888.5	86.1	3012	23	AA099289	Hepatitis C virus
36	885.5	85.8	3011	14	AA140120	HCV genomic amino
37	884.5	85.7	687	16	AA179223	pHCV150-encoded se
38	884.5	85.7	1648	16	AA179221	pHCV176-encoded se
39	884.5	85.7	1766	10	AA192041	Sequence encoded i
40	884.5	85.7	1786	10	AA190158	Protein sequence o
41	884.5	85.7	2261	10	AA190164	Peptide encoded by
42	884.5	85.7	2301	10	AA192047	Sequence encoded i
43	884.5	85.7	2436	10	AA192050	Sequence encoded i
44	884.5	85.7	2436	10	AA192058	Peptide encoded by
45	884.5	85.7	2772	21	AA18540	Protein encoded by

#### ALIGNMENTS

RESULT 1  
AA15222  
ID AA15222 standard; protein: 197 AA.  
XX  
AC AA15222;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #4.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI: 2000-465976/40.  
DR N-PSDB; AA173331.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 -  
 XX  
 XX Claim 23; Fig 14; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 XX Sequence 197 AA;

Query Match 100.0%; Score 1032; DB 21; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 6e-101;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAOOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAOOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60  
 QY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLTPTCTCGSSDLY 120  
 DB 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLTPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2  
 AAB15221  
 ID AAB15221 standard; protein; 197 AA.

XX AC AAB15221;  
 XX DT 19-DEC-2000 (first entry)  
 XX DE Hepatitis C virus NS4A-NS3 fusion protease #3.  
 XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX KW liver failure; liver cancer; mutant; mutein.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN WO200040707-A1.  
 XX PD 13-JUL-2000.  
 XX PF 06-JAN-2000; 2000WO-US00345.  
 XX PR 08-JAN-1999; 99US-0115271.  
 XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 XX DR N-PSDB; AAA73330.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 -  
 XX  
 XX Claim 23; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 XX Sequence 197 AA;

Query Match 98.4%; Score 1015; DB 21; Length 197;  
 Best Local Similarity 98.5%; Pred. No. 3.8e-99;  
 Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAOOTRGECCQETSGTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAOOTRGECCQETSGTGRDKNQVEGEVQIVSTAQTFLA 60  
 QY 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLTPTCTCGSSDLY 120  
 DB 61 TCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPQGSRSLTPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLGKSSGGPILCPAGHAGVIFRAAVCTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3  
 AAB15223  
 ID AAB15223 standard; protein; 197 AA.  
 XX AC AAB15223;  
 XX DT 19-DEC-2000 (first entry)  
 XX DE Hepatitis C virus NS4A-NS3 fusion protease #5.  
 XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX KW liver failure; liver cancer; mutant; mutein.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN WO200040707-A1.  
 XX PD 13-JUL-2000.  
 XX PF 06-JAN-2000; 2000WO-US00345.  
 XX PR 08-JAN-1999; 99US-0115271.  
 XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 XX DR N-PSDB; AAA73332.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
PS Claim 23; Fig 15; 66pp; English.  
XX The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
XX variant.  
XX  
SQ Sequence 197 AA;  
Query Match 97.1%; Score 1002; DB 21; Length 197;  
Best Local Similarity 98.5%; Pred. No. 9e-98;  
Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MKKGSVIVGRINLSGDTAYAAQQTGEGGCOETSGTRDKNQVEGEVQIVSTATQTFLA 60  
DB 1 MKKGSVIVGRINLSGDTAYAAQQTGEGGCOETSGTRDKNQVEGEVQIVSTATQTFLA 60  
QY 61 TCINGVCTVYHGAGTRTIIASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTCTCGSSDLY 120  
DB 61 TSINGVLTVYHGAGTRTIIASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 4  
AAB15224  
ID AAB15224 standard; protein; 197 AA.  
XX AAB15224;  
XX 19-DEC-2000 (first entry)  
XX Hepatitis C virus NS4A-NS3 fusion protease #6.  
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX Hepatitis C virus.  
OS Synthetic.  
XX WO200040707-A1.  
XX 13-JUL-2000.  
XX 06-JAN-2000; 2000WO-US00345.  
XX 08-JAN-1999; 99US-0115271.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI; 2000-465976/40.

DR N-PSDB; AAA73333.  
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
PS Claim 23; Fig 16; 66pp; English.  
XX The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
XX variant.  
XX  
SQ Sequence 197 AA;  
Query Match 95.9%; Score 990; DB 21; Length 197;  
Best Local Similarity 97.0%; Pred. No. 1.7e-96;  
Matches 191; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MKKGSVIVGRINLSGDTAYAAQQTGEGGCOETSGTRDKNQVEGEVQIVSTATQTFLA 60  
DB 1 MKKGSVIVGRINLSGDTAYAAQQTGEGGCOETSGTRDKNQVEGEVQIVSTATQTFLA 60  
QY 61 TCINGVCTVYHGAGTRTIIASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTCTCGSSDLY 120  
DB 61 TSINGVLTVYHGAGTRTIIASPKGPVTQMTYNDKDLVGMQAPQGSRSLSLTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGPLLCPCPAGHAGVIFRAAVCTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 5  
AAB15220  
ID AAB15220 standard; protein; 195 AA.  
XX AAB15220;  
XX 19-DEC-2000 (first entry)  
XX Hepatitis C virus NS4A-NS3 fusion protease #2.  
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX Hepatitis C virus.  
OS Synthetic.  
XX WO200040707-A1.  
XX 13-JUL-2000.  
XX 06-JAN-2000; 2000WO-US00345.  
XX 08-JAN-1999; 99US-0115271.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI; 2000-465976/40.

DR WPI; 2000-465976/40.  
 XX N-PSDB; AAA73329.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX

PS Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX

SQ Sequence 195 AA;

Query Match 95.1%; Score 981; DB 21; Length 195;  
 Best Local Similarity 96.4%; Pred. No. 1.5e-95;  
 Matches 190; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCOETSQTGRDNKQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRIVLNG--AYAAQTRGEGCOETSQTGRDNKQVEGEVQIVSTATQTFLA 58  
 QY 61 TCINGVCWTVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPCTCGSSDLY 120  
 DB 59 TCINGVCWTVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPCTCGSSDLY 118  
 QY 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 DB 119 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 6  
 AAB15225  
 ID AAB15225 standard; protein; 197 AA.

XX AAB15225;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #7.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200040707-A1.

XX 13-JUL-2000.

PD 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.  
 DR N-PSDB; AAA73334.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX

PS Claim 23; Fig 17; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX

SQ Sequence 197 AA;

Query Match 95.0%; Score 980; DB 21; Length 197;  
 Best Local Similarity 96.4%; Pred. No. 1.9e-95;  
 Matches 190; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCOETSQTGRDNKQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCOETSQTGRDNKQVEGEVQIVSTATQTFLA 60  
 QY 61 TCINGVCWTVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPCTCGSSDLY 120  
 DB 61 TSINGVLMTVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 DB 121 LVTRHADVIPVRRRGDSRGSLLSPRISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 7  
 AAB15226  
 ID AAB15226 standard; protein; 197 AA.

XX AAB15226;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #8.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

PN WO200040707-A1.

XX 13-JUL-2000.

PD 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.



PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR N-PSDB: AAA73335.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

PT amino acid, useful for screening inhibitors that may treat hepatitis C

PT -

XX Example 5; Fig 18; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0  
CC wild-type sequence.

XX Sequence 197 AA;

Query Match 94.6%; Score 976; DB 21; Length 197;

Best Local Similarity 95.9%; Pred. No. 5.1e-95;

Matches 189; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTATQTFIA 60

Db 1 MKKKGSVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTAAQTFLA 60

QY 61 TCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMWPAQPGSRSLTPTCTCGSSDLY 120

Db 61 TCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMWPAQPGSRSLTPTCTCGSSDLY 120

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

Db 121 LVTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

QY 181 AVDFIPVESLETMRSP 197

Db 181 AVDFIPVESLETMRSP 197

RESULT 8

AAB15212

ID AAB15212 standard; protein; 195 AA.

XX AAB15212;

AC AAB15212;

DT 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease #1.

DE Hepatitis C virus; single chain recombinant complex; linker;

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer.

XX Hepatitis C virus.

OS Synthetic.

XX WO2000040707-A1.

PN 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

DR N-PSDB: AAA73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT -

XX Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C  
CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
CC essential for the replication of the virus, acting to cleave its  
CC replicative proteins from the polyprotein produced from the HCV genome.  
CC Inhibitors of the two proteins should be effective as antiviral  
CC treatments of HCV infection. This is useful as HCV can lead to chronic  
CC liver disease such as cirrhosis, liver failure and liver cancer. The  
CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
CC proteins which can be used to identify inhibitors of this type, as well  
CC as enabling structural studies of the protease and protease:inhibitor  
CC complexes.

XX Sequence 195 AA;

Query Match 91.3%; Score 942; DB 21; Length 195;

Best Local Similarity 93.9%; Pred. No. 2e-91;

Matches 185; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

QY 1 MKKKGSVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTATQTFIA 60

Db 1 MKKKGSVIVGRINLSGDTAYAAQOTRGEGCGOETSGTGRDKKNQVEGEVOIVSTAAQTFLA 58

QY 61 TCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMWPAQPGSRSLTPTCTCGSSDLY 120

Db 59 TCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMWPAQPGSRSLTPTCTCGSSDLY 118

QY 121 LVTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAK 180

Db 119 LVTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCAGHAGVIFRAAVCTRGVAK 178

QY 181 AVDFIPVESLETMRSP 197

Db 179 AVDFIPVESLETMRSP 195

RESULT 9

AAY24943

ID AAY24943 standard; Protein; 665 AA.

XX AAY24943;

AC AAY24943;

DT 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex SEQ ID NO:14.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;

KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX hydrophobic domain; covalent complex; detection; inhibitor.

OS Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

PN 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.  
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WPI: 1999-385385/32.  
 XX New hepatitis C virus covalent complexes  
 XX Claim 6; Page 90-92; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.

XX Sequence 665 AA:

Query Match 88.3%; Score 911.5; DB 20; Length 665;  
 Best Local Similarity 86.7%; Pred. No. 1.9e-87;  
 Matches 170; Conservative 16; Mismatches 7; Indels 3; Gaps 1;  
 QY 5 GSVIVIGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVOIVSTATQTFLAT 61  
 DB 22 GSVIVIGRILSGSGSITAYSQOTRGLLGCKKTSITGRDKNOVEGEVOIVSTATQSFAT 81  
 QY 62 CINGVCWTVYHGAGTTRTIASPKGPVTQMTYNDKDLVGMQAPGQSRSLTPTCTGSSDLYL 121  
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPIQMTYNDQDLVGMQAPPGARSLTPTCTGSSDLYL 141  
 QY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKA 181  
 DB 142 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCSGHAGVIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETTMRSP 217

RESULT 10  
 AAY24947  
 ID AAY24947 standard; Protein; 665 AA.  
 XX AC AAY24947;  
 XX DT 07-SEP-1999 (first entry)  
 XX DE HCV NS4A-NS3 complex SEQ ID NO:18.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.  
 OS Synthetic.

XX WO9928482-A2.  
 XX PD 10-JUN-1999.  
 XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.

PA (SCHE ) SCHERING CORP.  
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WPI: 1999-385385/32.  
 XX New hepatitis C virus covalent complexes  
 XX Claim 6; Page 100-102; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.

XX Sequence 665 AA:

Query Match 88.0%; Score 908.5; DB 20; Length 665;  
 Best Local Similarity 86.2%; Pred. No. 3.9e-87;  
 Matches 169; Conservative 17; Mismatches 7; Indels 3; Gaps 1;  
 QY 5 GSVIVIGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVOIVSTATQTFLAT 61  
 DB 22 GSVIVIGRILSGSGSITAYSQOTRGLLGCKKTSITGRDKNOVEGEVOIVSTATQSFAT 81  
 QY 62 CINGVCWTVYHGAGTTRTIASPKGPVTQMTYNDKDLVGMQAPGQSRSLTPTCTGSSDLYL 121  
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPIQMTYNDQDLVGMQAPPGARSLTPTCTGSSDLYL 141  
 QY 122 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCAGHAGVIFRAAVCTRGVAKA 181  
 DB 142 VTRHADVIPVRRRGDSRGSLLSPRISYLGKSGSGPLLCPCSGHAGVIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETTMRSP 217

RESULT 11  
 AAY24941  
 ID AAY24941 standard; Protein; 665 AA.  
 XX AC AAY24941;  
 XX DT 07-SEP-1999 (first entry)  
 XX DE HCV NS4A-NS3 complex SEQ ID NO:12.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.  
 OS Synthetic.

XX WO9928482-A2.  
 XX PD 10-JUN-1999.  
 XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX DR WPI; 1999-385385/32.  
 XX PT New hepatitis C virus covalent complexes  
 XX PS Claim 6; Page 85-87; 21lpp; English.

XX CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.

XX SQ Sequence 665 AA;

Query Match 87.9%; Score 907.5; DB 20; Length 665;  
 Best Local Similarity 86.7%; Pred. No. 5e-87;  
 Matches 170; Conservative 15; Mismatches 8; Indels 3; Gaps 1;

QY 5 GSVIVTVGRINLGGD---TAYAQOTRGECCQTSOTGRKNOVEGEVOIVSTATOTFLAT 61  
 DB 22 GSVIVTVGRINLGGSGITAYISQOTRGLGCKITSLTGRKNOVEGEVOIVSTATOSFLAT 81  
 QY 62 CINGVCWTYVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121  
 DB 82 CVNGVCWTYVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSILPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCGAGHVGIFRAAVCTRGVAKA 181  
 DB 142 VTRHADVIPVRRRGDSRGLSPRPVSYLVKSGSGGPLLCPCSGHVGIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETMRSP 217

RESULT 12  
 AAY24942  
 ID AAY24942 standard; Protein; 665 AA.

XX AC AAY24942;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex SEQ ID NO:13.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX OS Hepatitis C virus.  
 OS Synthetic.

XX PN WO9928482-A2.

XX PD 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.

XX PR 28-JUL-1998; 98US-0094331.

XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE ) SCHERING CORP.

XX PI

PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX DR WPI; 1999-385385/32.  
 XX PT New hepatitis C virus covalent complexes  
 XX PS Claim 6; Page 88-90; 21lpp; English.

XX CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.

XX SQ Sequence 665 AA;

Query Match 87.9%; Score 907.5; DB 20; Length 665;  
 Best Local Similarity 86.7%; Pred. No. 5e-87;  
 Matches 170; Conservative 15; Mismatches 8; Indels 3; Gaps 1;

QY 5 GSVIVTVGRINLGGD---TAYAQOTRGECCQTSOTGRKNOVEGEVOIVSTATOTFLAT 61  
 DB 22 GSVIVTVGRINLGGSGITAYISQOTRGLGCKITSLTGRKNOVEGEVOIVSTATOSFLAT 81  
 QY 62 CINGVCWTYVYHGAGTRTIAASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121  
 DB 82 CVNGVCWTYVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSILPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRRGDSRGLSPRPISYLVKSGSGGPLLCPCGAGHVGIFRAAVCTRGVAKA 181  
 DB 142 VTRHADVIPVRRRGDSRGLSPRPVSYLVKSGSGGPLLCPCSGHVGIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETMRSP 217

RESULT 13  
 AAY17880  
 ID AAY17880 standard; Protein; 216 AA.

XX AC AAY17880;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex SEQ ID NO:4.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX OS Hepatitis C virus.  
 OS Synthetic.

XX PN WO9928482-A2.

XX PD 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.

XX PR 28-JUL-1998; 98US-0094331.

XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX DR WPI; 1999-385385/32.

XX PT New hepatitis C virus covalent complexes

XX PS Claim 6; Page 76-77; 21pp; English.

XX CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity. The helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.

XX SQ Sequence 216 AA;

Query Match 87.6%; Score 904.5; DB 20; Length 216;

Best Local Similarity 86.7%; Pred. No. 2.2e-87;

Matches 169; Conservative 16; Mismatches 7; Indels 3; Gaps 1;

OY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOFTSQTGRDNKQVEGEVQIYSTATQTFLAT 61

DB 22 GSVVIVGRILSGSGSITAYSQTRGLLGCKKTSLTGRDNKQVEGEVQVYSTATQSFAT 81

OY 62 CINGVCWTVYHGACRTIASPKGPVTOMYTNVDKLVGWQAPQGSRLTPTCTGSSDLYL 121

DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPQARSLTPTCTGSSDLYL 141

OY 122 VTRHADVIPVRRRGRSGSLSPRISYLGSGSGPILCPAGHAGVGFRAAVCTRGVAKA 181

DB 142 VTRHADVIPVRRRGRSGSLSPRPVSYLGSGSGPILCPSGHAGVGFRAAVCTRGVAKA 201

OY 182 VDFPVPESLETTMRS 196

DB 202 VDFPVPESMETTMRSP 216

RESULT 14

AY24945  
ID AAY24945 standard; Protein: 665 AA.

XX AC AAY24945;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex. SEQ ID NO:16.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX PN W09928482-A2.

XX PD 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.

XX PR 28-JUL-1998; 98US-0094331.

XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX DR

DR WPI; 1999-385385/32.

XX PT New hepatitis C virus covalent complexes

XX PS Claim 6; Page 95-97; 21pp; English.

XX CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity. The helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.

XX SQ Sequence 665 AA;

Query Match 87.6%; Score 904.5; DB 20; Length 665;

Best Local Similarity 86.2%; Pred. No. 1e-86;

Matches 169; Conservative 16; Mismatches 8; Indels 3; Gaps 1;

OY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOFTSQTGRDNKQVEGEVQIYSTATQTFLAT 61

DB 22 GSVVIVGRILSGSGSITAYSQTRGLLGCKKTSLTGRDNKQVEGEVQVYSTATQSFAT 81

OY 62 CINGVCWTVYHGACRTIASPKGPVTOMYTNVDKLVGWQAPQGSRLTPTCTGSSDLYL 121

DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGWQAPPQARSLTPTCTGSSDLYL 141

OY 122 VTRHADVIPVRRRGRSGSLSPRISYLGSGSGPILCPAGHAGVGFRAAVCTRGVAKA 181

DB 142 VTRHADVIPVRRRGRSGSLSPRPVSYLGSGSGPILCPSGHAGVGFRAAVCTRGVAKA 201

OY 182 VDFPVPESLETTMRS 197

DB 202 VDFPVPESMETTMRSP 217

RESULT 15

AY24946  
ID AAY24946 standard; Protein: 665 AA.

XX AC AAY24946;

XX DT 07-SEP-1999 (first entry)

XX DE HCV NS4A-NS3 complex. SEQ ID NO:17.

XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX PN W09928482-A2.

XX PD 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.

XX PR 28-JUL-1998; 98US-0094331.

XX PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX DR WPI; 1999-385385/32.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds  
(without alignments)  
949,524 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSWIVGRINLSGDTA.....VAKAVDFIPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	884.5	85.7	3011	1 POLG_HCV1	P26664 h genome po
2	878.5	85.1	3011	1 POLG_HCVH	P27958 h genome po
3	867.5	84.1	3010	1 POLG_HCVTW	P29846 h genome po
4	857.5	83.1	3010	1 POLG_HCVJT	Q00269 h genome po
5	853.5	82.7	3010	1 POLG_HCVBK	P26663 h genome po
6	853.5	82.7	3010	1 POLG_HCVJA	P26662 h genome po
7	677	65.6	3033	1 POLG_HCVJ6	P26660 h genome po
8	675	65.4	3033	1 POLG_HCVJ8	P26661 h genome po
9	86	8.3	321	1 HHOA_ARATH	Q9sel7 arabidopsis
10	84.5	8.2	485	1 Y136_TREPA	O83172 treponema p
11	81	7.8	452	1 AAMP_HUMAN	Q13685 homo sapien
12	78.5	7.6	437	1 DEGL_ARATH	O22609 arabidopsis
13	78	7.6	1705	1 PTPO_MOUSE	P70289 mus musculu
14	76.5	7.4	264	1 CTRL_HUMAN	P40313 homo sapien
15	76.5	7.4	323	1 VPRT_SMRVH	P21407 squirrel mo
16	76.5	7.4	333	1 MOSA_RHIME	Q07607 rhizobium m
17	76	7.4	911	1 TB11_NEIMB	Q09056 neisseria m
18	75.5	7.3	209	1 PAAD_PSEAE	O9hx08 pseudomonas
19	75.5	7.3	2663	1 CENE_HUMAN	Q02224 homo sapien
20	75	7.3	388	1 ODPT_HUMAN	P29803 homo sapien
21	75	7.3	455	1 TMS5_MOUSE	Q9er04 mus musculu
22	75	7.3	594	1 NIR_SPIOL	P05314 spinacia ol
23	74.5	7.2	706	1 TRFE_HORSE	P27425 equus caball
24	74.5	7.2	764	1 ICCR_DROME	O08180 drosophila
25	74	7.2	844	1 CN4A_RAT	P54748 rattus norv
26	73.5	7.1	263	1 GRAC_MOUSE	O35205 mus musculu
27	73	7.1	259	1 IBT1_HUMAN	P08833 homo sapien
28	72.5	7.0	452	1 MTLD_ECOLI	P23931 escherichia
29	72.5	7.0	2768	1 THYG_RAT	P06882 rattus norv
30	72	7.0	349	1 TRPD_PSEPU	P20575 pseudomonas
31	72	7.0	387	1 GALL_STRCO	Q9k3s8 streptomyce
32	72	7.0	1165	1 POL_GALV	P21414 gibbon ape
33	72	7.0	1210	1 EGFR_MOUSE	Q01279 mus musculu

## ALIGNMENTS

RESULT 1  
POLG\_HCV1

```

ID POLG_HCV1 STANDARD; PRT: 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RA *Genetic organization and diversity of the hepatitis C virus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62321; AAA45676.1; -.
CC PIR; A39166; GNMVC3.
CC PDB; 1AIV; 16-FEB-99.
CC PDB; 1HEI; 25-NOV-98.
CC MEROPS; S29.001; -.
CC MEROPS; O39.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

```

DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002186; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_pfsam.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RDRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PolyProtein: Glycoprotein; Transferase: RNA-directed RNA polymerase;  
 KW PolyProtein: Glycoprotein; Envelope protein; Helicase; ATP-binding;  
 KW Core protein; Coat protein; Nonstructural protein; Hydrolase; Serine protease;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).  
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NF\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;  
 Query Match 85.78; Score 884.5; DB 1; Length 3011;  
 Best Local Similarity 84.38; Pred. No. 4.3e-76;  
 Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;  
 3 KKGSWIVGRIN-----LSGDTAYAAQQTGRGCGCQTSOTGRKKNQVEGEIVST 53

Db 1005 RRGREILLGADGWSKGRLLAPITAYAAQQTGRGCGCQTSOTGRKKNQVEGEIVST 1064  
 QY 54 ATOTFLATCINGCVTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113  
 Db 1065 AAOTFLATCINGCVTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 1124  
 QY 114 CGSSDLYLVRHADVIPVRRGDSRGLSPISLYLKSGSGPILCPAGHAGVIGIFRAAV 173  
 Db 1125 CGSSDLYLVRHADVIPVRRGDSRGLSPISLYLKSGSGPILCPAGHAGVIGIFRAAV 1184  
 QY 174 CTRGAKAVDFIPVESLETMRSP 197  
 Db 1185 CTRGAKAVDFIPVESLETMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID POLG\_HCVH AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=111108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RL [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RT Nat. Struct. Biol. 4:463-467(1997).  
 RL [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVITY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M67463; AAA45534.1; -  
DR PIR: A36814; GNMVCH.  
DR PDB: 1HEI; 25-NOV-98.  
DR PDB: 1A1V; 16-FEB-99.  
DR PDB: 1A1R; 17-JUN-98.  
DR MEROPS: S29.001; -  
DR MEROPS: U39.001; -  
DR TRANSFAC: T04155; -  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRP.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; Helicase\_C; 1.  
DR Pfam: PF00998; Viral\_RdRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART: SM00487; bEXDC; 1.  
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure.  
KW INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.  
FT CHAIN 192 383 CAPSID PROTEIN C.  
FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.  
FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.  
FT CHAIN 810 1026 PROTEIN P7.  
FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.  
FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.  
FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4a.  
FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4b.  
FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5a.  
FT CHAIN 3011 369 POTENTIAL.  
FT TRANSMEM 347 369 POTENTIAL.  
FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
FT SITE 1316 1319 DECH BOX.  
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT 305 305 CARBOHYD  
FT 417 417 CARBOHYD  
FT 423 423 CARBOHYD  
FT 430 430 CARBOHYD  
FT 448 448 CARBOHYD  
FT 476 476 CARBOHYD  
FT 532 532 CARBOHYD  
FT 540 540 CARBOHYD  
FT 556 556 CARBOHYD  
FT 576 576 CARBOHYD  
FT 623 623 CARBOHYD  
FT 645 645 CARBOHYD  
FT STRAND 1224 1226  
FT TURN 1232 1233  
FT TURN 1236 1238  
FT HELIX 1239 1246  
FT TURN 1247 1248  
FT STRAND 1251 1255  
FT HELIX 1258 1271  
FT TURN 1272 1272  
FT STRAND 1277 1280  
FT TURN 1281 1282  
FT STRAND 1283 1285  
FT HELIX 1291 1295  
FT TURN 1302 1303  
FT STRAND 1312 1316  
FT TURN 1317 1319  
FT HELIX 1323 1335  
FT TURN 1336 1340  
FT STRAND 1343 1347  
FT TURN 1352 1353  
FT STRAND 1360 1361  
FT STRAND 1362 1366  
FT STRAND 1368 1368  
FT TURN 1373 1375  
FT TURN 1376 1377  
FT STRAND 1378 1380  
FT HELIX 1382 1385  
FT STRAND 1389 1393  
FT HELIX 1397 1409  
FT TURN 1410 1411  
FT STRAND 1414 1417  
FT TURN 1419 1420  
FT STRAND 1432 1436  
FT TURN 1438 1439  
FT STRAND 1450 1453  
FT STRAND 1456 1463  
FT STRAND 1471 1478  
FT STRAND 1480 1480  
FT HELIX 1481 1488  
FT TURN 1489 1490  
FT STRAND 1497 1501  
FT STRAND 1507 1507  
FT STRAND 1511 1511  
FT HELIX 1514 1527  
FT HELIX 1532 1544  
FT STRAND 1550 1550  
FT HELIX 1555 1564  
FT HELIX 1570 1578  
FT TURN 1579 1580  
FT HELIX 1584 1597  
FT TURN 1598 1598  
FT HELIX 1606 1611  
FT TURN 1614 1618  
FT STRAND 1622 1623  
FT STRAND 1627 1627  
FT STRAND 1635 1636  
FT HELIX 1640 1652  
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;  
Query Match 85.1%; Score 878.5; DB 1; Length 3011;  
Best Local Similarity 83.3%; Pred. No. 1.6e-75;







Db 1005 RRGREILLGPADSIQGGWRLAPITAYAAQTRCLLCIVTSLTRDKNQVEGEVQWST 1064  
QY 54 ATQTFELATCINGCWVTHGAGRTTASPGVPTOMYNTDKDLVGHQAPQGSRLTPCT 113  
Db 1065 ATQSFELATCINGCWVTHGAGRTTASPGVPTOMYNTDKDLVGHQAPQGSRLTPCT 1124  
QY 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGKSSGGLPLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRPISYLGKSSGGLPLCPAGHAGVIFRAAV 1184  
QY 174 CTRGVAKAVDFIPVESLETHMRSP 197  
Db 1185 CTRGVAKAVDFIPVESLETHMRSP 1208

RESULT 5  
POLG\_HCVBK  
ID POLG\_HCVBK STANDARD; PRT: 3010 AA.  
AC P26663;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (NS1); Envelope glycoprotein E3 (GP45) (NS2) (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (Isolate BK) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11105;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1847440;  
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,  
RA Onishi E., Andoh T., Yoshida I., Okayama H.;  
RT "Structure and organization of the hepatitis C virus genome isolated  
RT from human carriers.";  
RL J. Virol. 65:1105-1113(1991).  
RN [2]  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=96235224; PubMed=8647104;  
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;  
RT "Non-structural protein 3 of hepatitis C virus inhibits  
RT phosphorylation mediated by cAMP-dependent protein kinase.";  
RL Eur. J. Biochem. 237:611-618(1996).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.  
RX MEDLINE=97015088; PubMed=8861916;  
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,  
RA Moomaw E.W., Adachi T., Hostomsky Z.;  
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a  
RT trypsin-like fold and a structural zinc binding site.";  
RL Cell 87:331-342(1996).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.  
RX MEDLINE=98227846; PubMed=9568891;  
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,  
RA Steinkuehler C., Tomei L., De Francesco R., Kuo L.C., Chen Z.;  
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C  
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";  
RL Protein sci. 7:837-847(1998).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC {RNA}(N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M58335; AAA72945.1; -  
DR PIR: A38465; GNWVTC.  
DR PDB: 1A1Q; 25-MAR-98.  
DR PDB: 1JXP; 14-JAN-98.  
DR PDB: 1NS3; 08-APR-98.  
DR PDB: 1C2P; 15-NOV-00.  
DR PDB: 1CSJ; 08-NOV-99.  
DR PDB: 1GX5; 09-APR-02.  
DR PDB: 1GX6; 10-APR-02.  
DR PDB: 1QVY; 26-JUN-00.  
DR PDB: 80HM; 20-APR-99.  
DR MEROPS: S29.001; -  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV capsid.  
DR InterPro: IPR002521; HCV core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV NS1.  
DR InterPro: IPR002518; HCV NS2.  
DR InterPro: IPR004109; HCV NS3.  
DR InterPro: IPR000745; HCV NS4a.  
DR InterPro: IPR001490; HCV NS4b.  
DR InterPro: IPR002868; HCV NS5a.  
DR InterPro: IPR002166; HCV RdRP.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV capsid; 1.  
DR Pfam: PF01542; HCV core; 1.  
DR Pfam: PF01539; HCV env; 1.  
DR Pfam: PF01560; HCV NS1; 1.  
DR Pfam: PF01538; HCV NS2; 1.  
DR Pfam: PF02907; HCV NS3; 1.  
DR Pfam: PF01006; HCV NS4a; 1.  
DR Pfam: PF01001; HCV NS4b; 1.  
DR Pfam: PF01506; HCV NS5a; 1.  
DR Pfam: PF00998; Viral RdRP; 1.  
DR ProDom: PD186062; HCV NS1; 1.  
DR SMART: SM00487; DEXDC; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure.  
FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).  
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM.  
FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
FT SITE 1316 1319 DECH BOX.

CARBOHYD	196	196	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	209	209	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	234	234	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	250	250	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	305	305	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	417	417	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	423	423	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	430	430	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	448	448	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	532	532	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	540	540	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	556	556	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	576	576	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	623	623	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	645	645	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	2041	2041	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	2077	2077	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	2240	2240	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	2529	2529	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT	2788	2788	N-LINKED	(GLCNAC..)	(POTENTIAL..)
CARBOHYD	2798	2798	N-LINKED	(GLCNAC..)	(POTENTIAL..)
FT					

## RESULT 6

POLG\_HCVJA STANDARD; PRT; 3010 AA.

AC P26662;

AD 01-AUG-1992 (Rel. 23, Created)

AE 01-AUG-1992 (Rel. 23, Last sequence update)

AF 28-FEB-2003 (Rel. 41, Last annotation update)

AG Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

AH Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

AI (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

AJ (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

AK (EC 3.4.22.198); Nonstructural protein NS4A (P4); Nonstructural protein

AL NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

AM NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

AN Hepatitis C virus (isolate Japanese) (HCV).

AO Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

AP Hepacivirus

AQ NCBI\_TaxID=11116;

AX [1]

AY SEQUENCE FROM N.A.

BA MEDLINE-91088550; PubMed-2175903;

BB Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,

BC Sugimura T., Shimotohno K.;

BD "Molecular cloning of the human hepatitis C virus genome from

BE Japanese patients with non-A, non-B hepatitis.";

BF Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).

BG [2]

CH DISCUSSION OF SEQUENCE.

CI MEDLINE-91192160; PubMed-1849488;

CJ Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,

CK Ohkoshi S., Shimotohno K.;

CL "Molecular structure of the Japanese hepatitis C viral genome.";

CM FEBS Lett. 280:325-328(1991).

CN -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CO HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CP NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CQ -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CR precursor polyprotein, commonly with Asp or Glu in the P6

CS position, Cys or Thr in P1 and Ser or Ala in P1'.

CT -I- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

CU [RNA](N).

CV -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CW LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

DX PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

DY PROTEIN C AND MRNA.

EA -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

EB -----

EC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

ED between the Swiss Institute of Bioinformatics and the EMBL Outstation -

EE the European Bioinformatics Institute. There are no restrictions on its

EF use by non-profit institutions as long as its content is in no way

EG modified and this statement is not removed. Usage by and for commercial

EH entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EI -----

EJ EXEL; D90208; BAA14233.1; -

EL PIR; A39253; GNRVCJ.

EM HSSP; P26663; LUXP.

EN MEROPS; S29.001; -

EO MEROPS; U39.001; -

EP InterPro; IPR001410; DEAD.

EQ InterPro; IPR002522; HCV\_capsid.

ER InterPro; IPR002521; HCV\_core.

ES InterPro; IPR002519; HCV\_env.

ET InterPro; IPR002531; HCV\_NS1.

EU InterPro; IPR002518; HCV\_NS2.

EV InterPro; IPR004109; HCV\_NS3.

EW InterPro; IPR000745; HCV\_NS4a.

EX InterPro; IPR001490; HCV\_NS4b.

EY InterPro; IPR002868; HCV\_NS5a.

FA InterPro; IPR002166; HCV\_RdRp.



DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Coat protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 136 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA: 320165 MW: F957F5C1A273BE9E CRC64;  
 Query Match 65.6%; Score 677; DB 1; Length 3033;  
 Best Local Similarity 68.7%; Pred. No. 3e-56;  
 Matches 123; Conservative 26; Mismatches 30; Indels 0; Gaps 0;  
 QY 19 TAYAQTRGEECCQTSQTRDKNQVEGVQIVSTATQTFATCINGVGVYHGAQRT 78  
 1034 TAYAQTRGLLGTIVVSMTGRDKTEQAGEIQVLSVTQSFGLTTISVGLVWTVYHGAQNK 1093  
 QY 79 IASPGPVQMTVNDKDLVGNQAPQGSRLPTCTCGSSDLVLTVRHADVLPVRRGDSR 138  
 1094 LAGSRGPVQMTYSSAEGDLVGMPSPPGTKSLPECTCGAVDLYLVTRNADVPARRRGDKR 1153  
 QY 139 GSLLSPRIYILKSGSGGPLLCAGHANGVIFRAAVCTRGKAVADVIFVESLETMRSP 197  
 1154 GALLSPRLSTLKGSGGPEVLCPRGHAVGVFRAAVCSRGVAKSIDIFPVEILDIVTRSP 1212  
 Db 1154 GALLSPRLSTLKGSGGPEVLCPRGHAVGVFRAAVCSRGVAKSIDIFPVEILDIVTRSP 1212  
 RESULT 8  
 POLG\_HCVJ8  
 ID POLG\_HCVJ8 STANDARD; PRT: 3033 AA.  
 AC P26661;

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11115;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92230232; PubMed-1314459;  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishihiro S.;  
 RT \*Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes\*;  
 RL Virology 188:331-341(1992).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10988; BAA01761.1; -  
 CC PIR: A40250; GNMVJ8.  
 CC HSSP: P27958; 1HEI.  
 CC MEROPS: S29.001; -  
 CC MEROPS: U39.001; -  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV\_capsid.  
 CC InterPro: IPR002521; HCV\_core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC InterPro: IPR002531; HCV\_NS1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR004109; HCV\_NS3.  
 CC InterPro: IPR000745; HCV\_NS4a.  
 CC InterPro: IPR001490; HCV\_NS4b.  
 CC InterPro: IPR002868; HCV\_NS5a.  
 CC InterPro: IPR002166; HCV\_RdRp.  
 CC InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro: IPR007094; RNA\_pol\_PSVir.  
 CC Pfam: PF01543; HCV\_capsid; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01539; HCV\_env; 1.  
 CC Pfam: PF01560; HCV\_NS1; 1.  
 CC Pfam: PF01538; HCV\_NS2; 1.  
 CC Pfam: PF02907; HCV\_NS3; 1.  
 CC Pfam: PF01006; HCV\_NS4a; 1.  
 CC Pfam: PF01001; HCV\_NS4b; 1.  
 CC Pfam: PF01506; HCV\_NS5a; 1.  
 CC Pfam: PF00998; Viral\_RdRp; 1.

DR PRODom: PDI86062; HCV\_NSL; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;  
 Query Match 65.4%; Score 675; DB 1; Length 3033;  
 Best local similarity 69.3%; Pred. No. 4.6e-56;  
 Matches 124; Conservative 24; Mismatches 31; Indels 0; Gaps 0;  
 QY 19 TAYAQOTRCEGCQTSOTGRDKNOVEGVQIVSTATOTFLATCINGVCMVYHGAGT 78  
 Db 1034 TAYTOOTRGLLGAIVLSLGRDKNEAGOVQVLSVTOFLGTSIGVLYVYHGAGNT 1093  
 QY 79 IASPGPVQMTNVDKLVGQAPQGSRLTPTCGSSDLVLYTRADVPIVRRGDSR 138  
 Db 1094 LAGPGPVQMTNYSABGLVQWPPSPGKSLDPTCGGAVDLYLVTRADVPIVRRKDDR 1153  
 QY 139 GSLLSPRISYLSKSGGGLPCPAGHAYGIFRAVCTRGVAKAVDFIPVSELTMRSP 197  
 Db 1154 GALLSPRLSTLKGSGGGLVLCRSHAVGLFRFAVCAVGVAKSIDFIPVSLDVATTP 1212  
 RESULT 9  
 HHOA\_ARATH  
 ID HHOA\_ARATH STANDARD; PRT: 321 AA.  
 AC Q9SEL7: Q49507;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).  
 GN HHOA OR AT4G18370 OR F28J12.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucoids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RP SEQUENCE FROM N.A.  
 RN Lensch M.H.A., Sokolenko A., Herrmann R.G.;  
 RA "Identification and characterization of the chloroplast HhoA protease,  
 RT a homolog to the bacterial periplasmic protease HhoA";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RC MEDLINE=20083488; PubMed=10617198;  
 RX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesethoelt A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ausorge W., Brandt P., Grivell L., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delsen J., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weizenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Moijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berner S., Hempel S., Feldpausch M., Lanberth S., Van den Daele H.,  
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,  
 RA Neumann S., Argiriou A., Vitale D., Liquori R., Piravandi E.,  
 RA Massenat O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,  
 RA Perez-Perez A., Furnelle B., Bent E., Johnson S., Tacón D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Parnell P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Zaccaria L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Steneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Spieth J., Ryan E., Andrews S., Pepin K., Hillier L.,  
 RA Nelson J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,  
 RA Antonov B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:769-777(1999).  
 RN SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.  
 RP Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,  
 RA Kieselbach T.;  
 RT "The chloroplast lumen from Arabidopsis thaliana";  
 CC 1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.  
 CC 2- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
 CC 3- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
 CC gene model prediction. AT4G18370 and AT4G18375 were originally  
 CC fused into a single gene.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its







CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M95627; AAA68889.1; -  
DR PIR; I39383; I39383.  
DR Genew; HGNC:18; AAMP.  
DR MIM; 603488; -  
DR GO; 0008201; F:heparin binding activity; TAS.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 8.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD\_REPEATS.1; 1.  
DR PROSITE; PS0082; WD\_REPEATS.2; 6.  
DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
KW Repeat; WD repeat.  
FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).  
FT REPEAT 71 77 POLY-GLU.  
FT REPEAT 107 138 WD 1.  
FT REPEAT 150 180 WD 2.  
FT REPEAT 190 220 WD 3.  
FT REPEAT 231 261 WD 4.  
FT REPEAT 276 306 WD 5.  
FT REPEAT 333 363 WD 6.  
FT REPEAT 374 404 WD 7.  
FT REPEAT 416 446 WD 8.  
SQ SEQUENCE 452 AA; 49015 MW; DA1413D25EB236C0 CRC64;  
  
Query Match 7.8%; Score 81; DB 1; Length 452;  
Best Local Similarity 25.3%; Pred. No. 3;  
Matches 42; Conservative 13; Mismatches 47; Indels 54; Gaps 9;  
  
QY 68 VYVHGAGTRTIASPKGPVTQMTYVNDKLVGNOAPGGSRL-----TPCTCGSSDLV 122  
D 197 MNEWH-----PRAPVLLAGT-ADGNTWMKVPNGDKCTFGPNCPTACGR----- 240  
QY 123 TRHADVTPVRR-----GDSRGS-----LLSPRTSYLKGSSG--GPLLCPA----- 162  
D 241 -----VLPDGKRAVGVYEDGTRIRLDKQSPHVLKGTGCHGGLTCVAANQDGSLLT 295  
QY 163 -----GHAVGIFR-----AAVCTRGVAKAVDFIPVESL 190  
D 296 GSVDCQAKLVSAITGKVVGVPERPETAQSPSLGEGEESNSVESL 341  
  
RESULT 12  
DEGLARATH  
ID DEGLARATH STANDARD; PRT; 437 AA.  
AC 022609; O9LK85;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).  
GN DEGP1 OR DEGP OR Ar3G27925 OR K16N12.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=98175982; PubMed=9507020;  
RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;  
RT Identification and characterization of DegP, a serine protease  
RL associated with the luminal side of the thylakoid membrane.;  
RN J. Biol. Chem. 273:7094-7098(1998).  
RP SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RC

RX MEDLINE=20363099; PubMed=10907853;  
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
RT Structural analysis of Arabidopsis thaliana chromosome 3. II.  
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC  
RT and BAC clones.;  
RL DNA Res. 7:217-221(2000).  
RN [3]  
RP SEQUENCE OF 104-118.  
RC STRAIN=cv. Columbia;  
RA Kieselbach T., Bystedt M., Schroeder W.P.;  
RL Submitted (JUL-2000) to the SWISS-PROT data bank.  
CC !- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.  
CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,  
CC CAN DEGRADE BETA-CASEIN  
CC !- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND  
CC O-PHENANTHROLINE.  
CC !- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID  
CC MEMBRANE.  
CC !- INDUCTION: BY heat shock.  
CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
CC !- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF028842; AAC39436.1; -  
DR EMBL; AP000371; BAB02539.1; -  
DR EMBL; AP001302; BAB02539.1; JOINED.  
DR MEROPS; S01.279; -  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 1.  
DR PROSITE; PS0106; PDZ; 1.  
KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.  
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).  
FT CHAIN 104 437 PROTEASE DO-LIKE 1.  
FT DOMAIN 152 321 SERINE PROTEASE.  
FT DOMAIN 324 421 PDZ.  
FT ACT\_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).  
FT CONFLICT 12 23 HSPSSQLSNST -> SSTFLHSPSSHL (IN REF.  
FT CONFLICT 36 36 V -> I (IN REF. 2).  
FT CONFLICT 54 54 P -> S (IN REF. 2).  
FT CONFLICT 60 60 G -> R (IN REF. 2).  
FT CONFLICT 64 64 G -> D (IN REF. 2).  
FT CONFLICT 68 69 LL -> HF (IN REF. 2).  
FT CONFLICT 355 355 L -> V (IN REF. 2).  
FT CONFLICT 381 381 I -> V (IN REF. 2).  
FT CONFLICT 416 416 Q -> E (IN REF. 2).  
SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;  
  
Query Match 7.6%; Score 78.5; DB 1; Length 437;  
Best Local Similarity 25.6%; Pred. No. 5;  
Matches 44; Conservative 17; Mismatches 56; Indels 55; Gaps 7;  
  
QY 70 VYHAGTRTIASPKGPVTQMY-----TNVDKDLVGW-----QA 102  
D 150 VPQSGSGFVMDKQGHIVTYHVRGASDLRVLADQTTFDKAVGVDFDQDKDAVLRIDA 209  
QY 103 PQGSRSLTPTCGSSDLV-----TRHADVIPVRRGDSGSLSPRI 147  
RC

Db 210 PK--NKLRIPIVGVADLLVGVKQVFAIGNPGLDHTLTITGVISLRRREIS--SAATGRPI 265  
 Qy 148 SYL-----KSGSGGLPCPAGHAVGIFRAAVCTRGVAKAVDF-IPVESL 190  
 Db 266 QDVLTQDAAINPGSGGLDSSLTGLINTAIYSPSGASSGVGFSIPVDTV 317

RESULT 13  
 PTPO\_MOUSE STANDARD; PRT; 1705 AA.  
 AC P70289;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Embryonic stem cell protein tyrosine phosphatase precursor  
 DE (EC 3.1.3.48) (ES cell phosphatase).  
 GN ESP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic stem cells;  
 RX MEDLINE=97109513; PubMed=8951793;  
 RA Lee K., Nichols J., Smith A.;  
 RT Identification of a developmentally regulated protein tyrosine  
 RT phosphatase in embryonic stem cells that is a marker of  
 RT pluripotential epiblast and early mesoderm.\*;  
 RL Mech. Dev. 59:153-164(1996).  
 RN [2]  
 RP ERRATUM.  
 RA Lee K., Nichols J., Smith A.;  
 RL Mech. Dev. 61:213-215(1996).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE MAINTENANCE OF PLURIPOTENCY.  
 CC -!- DOWN-REGULATED DURING DIFFERENTIATION.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: DETECTABLE IN THE EPIBLAST OF OOCYTES AND  
 CC THROUGHOUT EARLY MOUSE EMBRYO DEVELOPMENT. IN ADULT, EXPRESSION IS  
 CC LOCALIZED IN GONADAL GERM CELLS.  
 CC -!- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.  
 CC -!- SIMILARITY: Contains 10 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 DR EMBL: U36488; AAC52868.1; -;  
 DR HSPB; P18052; IYFO.  
 DR MGD; MGI:108027; ESP.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_Pp.  
 DR Pfam; PF00041; fn3; 7.  
 DR Pfam; PF00102; Y\_phosphatase; 1.  
 DR PRINTS; PR00700; PRTYPHPTASE.  
 DR SMART; SM00060; FN3; 8.  
 DR SMART; SM00194; PTPC; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE; Repeat; Signal; Glycoprotein.  
 KW Hydrolase; Transmembrane; Repeat; Signal; Glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 1705 EMBRYONIC STEM CELL PROTEIN TYROSINE  
 FT PHOSPHATASE.  
 FT DOMAIN 19 1077 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1078 1100 POTENTIAL.

FT DOMAIN 1101 1705 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 35 110 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 126 199 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 216 288 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 304 373 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 393 454 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 471 543 FIBRONECTIN TYPE-III 6.  
 FT DOMAIN 563 634 FIBRONECTIN TYPE-III 7.  
 FT DOMAIN 657 722 FIBRONECTIN TYPE-III 8.  
 FT DOMAIN 742 813 FIBRONECTIN TYPE-III 9.  
 FT DOMAIN 831 905 FIBRONECTIN TYPE-III 10.  
 FT DOMAIN 1150 1418 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 1469 1700 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 1350 1350 BY SIMILARITY.  
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 551 551 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 570 570 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 970 970 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 982 982 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1705 AA; 186795 MW; 2783755F15387D5B CRC64;

Query Match 7.68; Score 78; DB 1; Length 1705;  
 Best Local Similarity 25.5%; Pred. No. 28;  
 Matches 42; Conservative 12; Mismatches 53; Indels 58; Gaps 8;

QY 9 IVGRINLSGDTAVAAQOTRG-EGCOETSTQTR-----DKNOVEGEVQIVSTATQT 57  
 Db 382 VEGSIWLAEASNAARMEVPVGARLWLEGLKATQGRALLYSVDAPGLGNISVSSGATHV 441  
 QY 58 FLATCINGVCWTVYHGAGTRTASPKGPVQMT-----NVDKDL-VGM 100  
 Db 442 TFCGLVPGAHYRV-----DIASSMGDITQSLTGYTSPLOSLEIIRNSPSDLTICW 494  
 QY 101 -QAQGSRSITPTCTCGSSDLYLVTRADVIPVRRGDSRGSLSP 144  
 Db 495 APAP-----GOMEGYKVTWHQD-----GSQRSP 517

RESULT 14  
 CTRL\_HUMAN STANDARD; PRT; 264 AA.  
 AC P40313;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chymotrypsin-like protease CTRL-1 precursor (EC 3.4.21.-).  
 GN CTRL OR CTRL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94093544; PubMed=8268911;  
 RA Larsen F., Solheim J., Kristensen T., Kolsto A.B., Prydz H.;  
 RT "A tight cluster of five unrelated human genes on chromosome  
 RT 16q22.1".  
 RL Hum. Mol. Genet. 2:1589-1595(1993).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; X71874; CAAS0710.1; -
CC EMBL; X71877; CAAS0711.1; -
CC PIR; I38136; I38136.
CC HSSP; P00763; IDPO.
CC MEROPS; S01.256; -
CC Genew; HGNC:2524; CTRL.
CC MIM; 118888; -
CC GO; GO:0005615; C:extracellular space; TAS.
CC GO; GO:0007386; P:digestion; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser.protease_Try.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PSS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 18 POTENTIAL
FT PROPEP 19 33 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 34 264 CHYMOTRYPSIN-LIKE PROTEASE CTRL-1.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOXYD 114 114 N-LINKED (GLNAC. . .) (POTENTIAL).
FT DISULFID 19 141 BY SIMILARITY.
FT DISULFID 60 76 HY SIMILARITY.
FT DISULFID 155 220 BY SIMILARITY.
FT DISULFID 187 201 BY SIMILARITY.
FT DISULFID 210 239 BY SIMILARITY.
SQ SEQUENCE 264 AA; 28002 MW; 3F629F02FA6DDB4 CRC64;

Query Match 7.4%; Score 76.5; DB 1; Length 264;
Best Local Similarity 25.9%; Pred. No. 4.3;
Matches 35; Conservative 18; Mismatches 51; Indels 31; Gaps 5;

QY 44 VEVEGVIVSTATQTFLATCINGCVWTVYHGAGRTTASPKGPVTVQMYTNVYDKDLVGVQAP 103
DB 118 MNDVTLKLLASPAQYTRISPC-----LASSNEALTEGLTCV---TTGWREL 163

QY 104 QGSRSLTPCTCGSSDLYLVYTRHADVIPVRRRGDSRGLSLSPRI-----SYLKGSSGG 156
DB 164 SGVGNVTPAHLOQVALPLVT-----VNCROYWGSSITDSMICAGGAGASSCQDSDGG 216

QY 157 PLLCPAGHA---VGI 168
DB 217 PLVCQKGNVWVLI 231

RESULT 15
VPRT_SMRVH STANDARD; PRT; 323 AA.
AC P21407;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease (EC 3.4.23.-).
GN PRT.
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).
CC Viruses; Retroid viruses; Retroviridae; Betaretrovirus.
OX NCBI_Taxid=11856;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE-95073750; PubMed-3201749;
RA Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K.,
RA Mitsunobu F.;
RT "Molecular cloning, complete nucleotide sequence, and gene structure
RT of the provirus genome of a retrovirus produced in a human
RT lymphoblastoid cell line.";
RL Virology 167:468-476(1988).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2.
CC -1- SIMILARITY: Contains 1 G-patch domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; M23385; AAA66452.1; ALT_INIT.
CC PIR; B31827; PRLJHD.
CC HSSP; P06968; IEDU.
CC MEROPS; A02.0PM; -
CC InterPro; IPR001995; Aspprotease_rtrv.
CC InterPro; IPR001969; Aspprotease_site.
CC InterPro; IPR001428; DeoxyUTPase.
CC InterPro; IPR000467; G_patch.
CC Pfam; PF00692; dUTPase; 1.
CC Pfam; PF01585; G_patch; 1.
CC Pfam; PF00077; rvp; 1.
CC SMART; SM00443; G_patch; 1.
CC PROSITE; PS00141; ASP_PROTEASE; 1.
CC PROSITE; PS00175; ASP_PROT_RETROV; 1.
CC PROSITE; PS00174; G_PATCH; 1.
CC Hydrolase; Aspartyl. protease.
FT DOMAIN 275 321 G-PATCH.
FT ACT_SITE 193 193 BY SIMILARITY.
SQ SEQUENCE 323 AA; 35126 MW; 5D6CEA38BA932786 CRC64;

Query Match 7.4%; Score 76.5; DB 1; Length 323;
Best Local Similarity 23.3%; Pred. No. 5.5;
Matches 34; Conservative 16; Mismatches 49; Indels 47; Gaps 5;

QY 42 NOVGEVOIVSTATQTFLATCINGCVWTVYHGAGRTTASPKG-----PVTOMYTN 92
DB 111 NDFEGEIHILSTKDL-----VTIPKGTFLAQIVILPLQQINSN 150

QY 93 VDKDLVGVQAPQSGSRSLTPCTCGSSDLYLV---TRHADVIPVRRRGDSRGLSL---SPR 145
DB 151 PHKPYRGASAP-----GSSDVYVWQQISQQRPTIKLKLNGKLFSGILDTGADAT 199

QY 146 PTSYLKSSGGPILCPAGHAVGIPRA 171
DB 200 VISYTHWPNWPLTTVATHLRGIGQA 225

Search completed: August 30, 2003, 19:13:46
Job time : 10.7567 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 : Search time 37.5921 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-16  
Perfect score: 1032  
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	902.5	87.5	4040	12 Q9IFH8	Q9ifh8 mucosal dis
2	888.5	86.1	3011	12 Q36579	Q36579 hepatitis c
3	884.5	85.7	2436	12 Q81756	Q81756 hepatitis c
4	884.5	85.7	3011	12 Q9IFB5	Q9ifb5 hepatitis c
5	884.5	85.7	3011	12 Q9EL58	Q9el58 hepatitis c
6	883.5	85.6	3011	12 Q3463	Q3463 hepatitis c
7	881.5	85.4	3011	12 Q36608	Q36608 hepatitis c
8	881.5	85.4	3015	12 Q9PW45	Q9pw45 hepatitis c
9	881.5	85.4	3015	12 Q9PW09	Q9pw09 hepatitis c
10	879	85.2	181	12 Q91RR8	Q91rr8 hepatitis c
11	879	85.2	181	12 Q91RT5	Q91rt5 hepatitis c
12	877	85.0	181	12 Q91RR5	Q91rr5 hepatitis c
13	877	85.0	181	12 Q91RR2	Q91rr2 hepatitis c
14	877	85.0	181	12 Q91RR9	Q91rr9 hepatitis c
15	876	84.9	181	12 Q91RR3	Q91rr3 hepatitis c
16	876	84.9	181	12 Q91RR4	Q91rr4 hepatitis c

17	876	84.9	181	12 Q91RS1	Q91rs1 hepatitis c
18	876	84.9	181	12 Q91RQ8	Q91rq8 hepatitis c
19	876	84.9	181	12 Q91RT1	Q91rt1 hepatitis c
20	876	84.9	181	12 Q91RR0	Q91rr0 hepatitis c
21	875.5	84.8	3011	12 Q36609	Q36609 hepatitis c
22	874	84.7	181	12 Q91RR6	Q91rr6 hepatitis c
23	874	84.7	181	12 Q91RS9	Q91rs9 hepatitis c
24	873	84.6	181	12 Q91RS3	Q91rs3 hepatitis c
25	872.5	84.5	3011	12 Q9DIT6	Q9dit6 hepatitis c
26	872	84.5	181	12 Q91RT4	Q91rt4 hepatitis c
27	872	84.5	181	12 Q91RS8	Q91rs8 hepatitis c
28	872	84.5	181	12 Q91RT3	Q91rt3 hepatitis c
29	872	84.5	181	12 Q91RS5	Q91rs5 hepatitis c
30	872	84.5	181	12 Q91RS7	Q91rs7 hepatitis c
31	872	84.5	181	12 Q91RT0	Q91rt0 hepatitis c
32	872	84.5	181	12 Q91RS2	Q91rs2 hepatitis c
33	871	84.4	181	12 Q91RS6	Q91rs6 hepatitis c
34	870.5	84.4	3010	12 Q9QP61	Q9qp61 hepatitis c
35	870	84.3	181	12 Q91RS4	Q91rs4 hepatitis c
36	869.5	84.3	3010	12 Q68533	Q68533 hepatitis c
37	869	84.2	181	12 Q91RR7	Q91rr7 hepatitis c
38	869	84.2	181	12 Q91RT6	Q91rt6 hepatitis c
39	869	84.2	3011	12 Q36610	Q36610 hepatitis c
40	868.5	84.2	361	12 Q70817	Q70817 hepatitis c
41	868	84.1	181	12 Q91RT8	Q91rt8 hepatitis c
42	867.5	84.1	361	12 Q70818	Q70818 hepatitis c
43	867	84.0	181	12 Q91RR9	Q91rr9 hepatitis c
44	866.5	84.0	3010	12 Q9DTE2	Q9dte2 hepatitis c
45	866.5	84.0	3010	12 Q99AU2	Q99au2 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q9IFH8	PRELIMINARY;	PRT; 4040 AA.
AC	Q9IFH8;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;		
OC	Pestivirus.		
OX	NCBI_TaxID=11099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20323484; PubMed=10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	"Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.;"		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF268278; AAF82566.1; "		
DR	HSSP; P26663; 1JXP.		
DR	MEROPS; S31.001; -.		
DR	InterPro; IPR000280; CDvir_endptsep80.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR001005; Myb_DNA_binding.		
DR	InterPro; IPR001568; RNase_T2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSVir.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF00271; Helicase_C; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		

```
DR PRINTS: PR00729; COVENDOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS05007; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
DR PROSITE: PS05031; RNASE_T2.2; 1.
KW ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein.
KW RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA; 453073 MW; ADS87791D055B9DC CRC64;

Query Match      87.5%; Score 902.5; DB 12; Length 4040;
Best Local Similarity 90.8%; Pred. No. 3e-81;
Matches 177; Conservative 5; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVIVIGRLNSGD---TAYAQOTRGECCQETSGTGRDNQVEGEVOIVSTATQTFLAT 61
DB 10 GSVIVIGRLVSGSITCAQOTRGLLCKIITSLTGRDNQVEGEVOIVSTATQTFLAT 69
QY 62 CINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL 121
DB 70 CINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCTCGSSDLYL 129
QY 122 VTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILLCPAGHAGVIFRAAVCTRGVAKA 181
DB 130 VTRHANVIPVRRGDSRGLSPRPISYLGSSGGPILLCPAGHAGVIFRAAVCTRGVAKA 189
QY 182 VDFIPVESLETTMRS 196
DB 190 VDFIPVENLETTMRS 204

RESULT 2
O36579 ID O36579 PRELIMINARY; PRT: 3011 AA.
AC O36579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF009606; AAB66324.1; -.
DR HSSP: P27958; 1HEY.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR000410; HCV_NS5a.
DR InterPro: IPR001490; HCV_NS5b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR001490; HCV_NS5b.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.

DR PRINTS: PF01543; HCV_capsid; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01560; HCV_NSI; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF02907; HCV_NS3; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF00271; Helicase_C; 1.
DR PFAM: PF00998; Viral_RdRP; 1.
DR PRODOM: PDI86062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05007; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EBE09C63C1B9 CRC64;

Query Match      86.1%; Score 888.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 5.4e-80;
Matches 172; Conservative 10; Mismatches 13; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGECCQETSGTGRDNQVEGEVOIVST 53
DB 1005 RKGQILIGPADGMVSKWRLIAPITAYAQOTRGLLCKIITSLTGRDNQVEGEVOIVST 1064
QY 54 ATQTFLATCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1065 ATQTFLATCINGVCWTVYHGAGTRTITASPKGPVTOMYTNVDQDLVGMQAPQGSRLTPTCT 1124
QY 114 CGSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILLCPAGHAGVIFRAAV 173
DB 1125 CGSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPILLCPAGHAGVIFRAAV 1184
QY 174 CTRGVAKAVDFIPVESLETTMRS 197
DB 1185 CTRGVAKAVDFIPVENLETTMRS 1208

RESULT 3
Q81756 ID Q81756 PRELIMINARY; PRT: 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL: M32084; AAA45677.1; -.
DR HSSP: P27958; 1A1V.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS5a.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR PFAM: PF01560; HCV_NSI; 1.
DR PFAM: PF01538; HCV_NS2; 1.
```

```

DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE; PS050507; RDRP_POSITIVE; 1.
DR PROSITE; PS050521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT NON_TER 2436
SQ SEQUENCE 2436 AA: 264734 MW: D7B9872900BE3125 CRC64;

Query Match 85.7%; Score 884.5; DB 12; Length 2436;
Best Local Similarity 84.3%; Pred. No. 1e-79;
Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVTVIGRIN-----LSGDTAYAAQOTRGECCQETSOTGRDKNOVEGEVQIVST 53
Db 555 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 614

QY 54 ATOTFLATCINGVCWTVYHGAGTRTASPKGPVQMTYNTVDKLVGQAPQGSRSILTPCT 113
Db 615 AAOTFLATCINGVCWTVYHGAGTRTASPKGPVQMTYNTVDKLVGQAPQGSRSILTPCT 674

QY 114 CGSSDLVLTTRHADVIPVRRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 173
Db 675 CGSSDLVLTTRHADVIPVRRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 734

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 735 CTRGVAKAVDFIPVENLETTMRSP 758

RESULT 4
Q91FE5 PRELIMINARY; PRT; 3011 AA.
ID Q91FE5
AC Q91FE5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
RC [-] SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1A1V.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.

```

```

DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS050507; RDRP_POSITIVE; 1.
DR PROSITE; PS050521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA: 327124 MW: 2489CE74AC864E58 CRC64;

Query Match 85.7%; Score 884.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.4e-79;
Matches 172; Conservative 9; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVTVIGRIN-----LSGDTAYAAQOTRGECCQETSOTGRDKNOVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGRDKNOVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTRTASPKGPVQMTYNTVDKLVGQAPQGSRSILTPCT 113
Db 1065 AAOTFLATCINGVCWTVYHGAGTRTASPKGPVQMTYNTVDKLVGQAPQGSRSILTPCT 1124

QY 114 CGSSDLVLTTRHADVIPVRRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 173
Db 1125 CGSSDLVLTTRHADVIPVRRRGDSRGLSPRISYLYKSGSGGPLLCPAGHAGVIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 5
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
ID Q9ELS8
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC [-] SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.

```



```
Query Match      85.6%; Score 883.5; DB 12; Length 3011;
Best Local Similarity 84.3%; Pred. No. 1.7e-79;
Matches 172; Conservative 8; Mismatches 15; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEGCOETSGTGRDKNQVEGEVQIVST 53
Db 1005 RKGREILLGPADGWSKGRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOYTNVDKDLVGWQAPQGSRLTPTCT 113
Db 1065 AATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMTNVQDLVGWPAQGSRLTPTCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLKSGSGGPLLCPCAGHAVGIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLKSGSGGPLLCPCAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608 PRELIMINARY; PRT: 3011 AA.
AC O36608;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF011751; AAB67036.1; -;
DR HSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR002516; HCV_NS3.
DR InterPro: IPR004109; HCV_NS4a.
DR InterPro: IPR000745; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PS_vir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.

DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDc; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match      85.4%; Score 881.5; DB 12; Length 3011;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEGCOETSGTGRDKNQVEGEVQIVST 53
Db 1005 RRGQEILLGPADGWSKGRLLAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOYTNVDKDLVGWQAPQGSRLTPTCT 113
Db 1065 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMTNVQDLVGWPAQGSRLTPTCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLKSGSGGPLLCPCAGHAVGIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGRDSRGLSPRISYLKSGSGGPLLCPCAGHAVGIFRAAV 1184

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 8
O9PWK5 PRELIMINARY; PRT: 3015 AA.
AC O9PWK5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177040; AAF01182.1; -;
DR EMBL: AF177038; AAF01180.1; -;
DR HSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
```



```

DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01539; HCV core; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR Pfam: PF0186062; DEXDC; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; 87D3BC1F190663A CRC64;

Query Match 85.4%; Score 881.5; DB 12; Length 3015;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQTRGEGCOETSTGRDNQVEGEVQIVST 53
DB 1009 RRGQELLGPADGMVSKGWRLLAPITAYAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1069 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGMQAPQGSRLTPTCT 1128

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 173
DB 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 1188

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRMRSP 1212

RESULT 9
Q9PMU9 PRELIMINARY; PRT; 3015 AA.
AC Q9PMU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99420396; PubMed-10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT *Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.;
RN [2]
RL Virology 262:250-263(1999).
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_dec.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV capsid; 1.
DR Pfam: PF01542; HCV core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR Pfam: PF0186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 85.4%; Score 881.5; DB 12; Length 3015;
Best Local Similarity 83.8%; Pred. No. 2.7e-79;
Matches 171; Conservative 10; Mismatches 14; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQTRGEGCOETSTGRDNQVEGEVQIVST 53
DB 1009 RRGQELLGPADGMVSKGWRLLAPITAYAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB 1069 ATOTFLATCINGVCWTVYHGAGTRTIASPKGPVIQMYTNVDQDLVGMQAPQGSRLTPTCT 1128

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 173
DB 1129 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLGSSGGPLLCPCPAGHAGVIFRAAV 1188

QY 174 CTRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTRMRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETHRSP 197

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US0965594/runat\_29082003\_151919\_28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biom62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0965594.ecgn\_1\_14686 -runat\_29082003\_151919\_28310 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pin:\*  
35: em.htg\_rtd:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	89.4	12734	6	ARI79057 Sequence
2	911.5	88.3	1398	6	ARI45264 Sequence
3	908.5	88.0	1998	6	ARI45268 Sequence
4	907.5	87.9	1998	6	ARI45262 Sequence
5	907.5	87.9	1998	6	ARI45263 Sequence
6	904.5	87.6	651	6	ARI45254 Sequence
7	904.5	87.6	1998	6	ARI45266 Sequence
8	904.5	87.6	1998	6	ARI45267 Sequence
9	903.5	87.5	1998	6	ARI45261 Sequence
10	903.5	87.5	2016	6	ARI45269 Sequence
11	902.5	87.5	12734	14	AF268278 Pestivirus
12	901.5	87.4	651	6	ARI45258 Sequence
13	900.5	87.3	651	6	ARI45252 Sequence
14	900.5	87.3	651	6	ARI45253 Sequence
15	900.5	87.3	1998	6	ARI45265 Sequence
16	900.5	87.3	2016	6	ARI45270 Sequence
17	900	87.2	648	6	ARI45274 Sequence
18	898	87.0	648	6	ARI45272 Sequence
19	897.5	87.0	651	6	ARI45256 Sequence
20	897.5	87.0	651	6	ARI45257 Sequence
21	897.5	87.0	651	6	ARI45260 Sequence
22	896.5	86.9	651	6	ARI45251 Sequence
23	896	86.8	648	6	ARI45273 Sequence
24	894	86.6	648	6	ARI45271 Sequence
25	893.5	86.6	651	6	ARI45255 Sequence
26	893.5	86.6	651	6	ARI45259 Sequence
27	891	86.3	8157	6	ARI27810 Sequence
28	891	86.3	8157	6	BD081911 Hepatitis
29	889	86.1	1932	6	ARI27809 Sequence
30	889	86.1	1932	6	BD081910 Hepatitis
31	888.5	86.1	9646	6	ARI10828 Sequence
32	888.5	86.1	9646	6	BD069982 Functiona
33	888.5	86.1	9646	14	AF009606 Hepatitis
34	888.5	86.1	12980	6	ARI10831 Sequence
35	888.5	86.1	12980	6	BD069985 Functiona
36	884.5	85.7	5360	6	ARI18686 Sequence
37	884.5	85.7	5360	6	I06434 Sequence 48
38	884.5	85.7	5360	6	I09328 Sequence 8
39	884.5	85.7	6785	6	ARI18692 Sequence
40	884.5	85.7	6785	6	I06440 Sequence 54
41	884.5	85.7	6785	6	I09329 Sequence 10
42	884.5	85.7	7310	6	ARI18696 Sequence
43	884.5	85.7	7310	6	I09331 Sequence 15
44	884.5	85.7	7310	14	HPCPOLYP M32084 Hepatitis C
45	884.5	85.7	8316	6	ARI18703 Sequence

# ALIGNMENTS

RESULT 1

ARL179057  
 LOCUS ARL179057 12734 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 1 from patent US 6326137.  
 ACCESSION ARL179057  
 VERSION ARL179057.1 GI:20220612  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 12734)  
 AUTHORS Hong Z., Lai V.C.H. and Lau J.Y.N.  
 TITLE Hepatitis C virus protease-dependent chimeric pestivirus  
 JOURNAL Patent: US 6326137-A 1 04-DEC-2001;  
 FEATURES  
 source Location/Qualifiers  
 1..12734 /organism="unknown"  
 BASE COUNT 4032 a 2604 c 3295 g 2803 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8,58e-67 Length: 12734  
 Score: 922.50 Matches: 180  
 Percent Similarity: 94.36% Conservative: 4  
 Best Local Similarity: 92.31% Mismatches: 8  
 Query Match: 89.39% Indels: 3  
 DB: 6 Gaps: 1  
 US-09-965-594-16 (1-197) x ARL179057 (1-12734)  
 QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 413 GGTAGTGTGTATTGTTGGTAGAATATGTTTATCTGTTAGTGTAGTACAGCGGGGTAC 472  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
 DB 473 GCCCAGACAGACAGAGCGCTCTAGGTGTAGAGTACCACTGTGCTGCTGCGCGGACAA 532  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 533 AACCAAGTGGAGGGTGAGTGCAGATCGTGTCAACTGCTACCCAAACCTTCCCTGGCAACG 592  
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 593 TGCATCAATGGGTATGCTGACATGCTTACACCGGGCGGACGAGGACCATCGCATCA 652  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 653 CCCAAGGCTCTGTCATCCAGATGTATACCAATGTGCACCAAGACCTTGTGGCTGSCCC 712  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 713 GCTCCTCAAGGTTCGGGTCTATTGACACCTGACACCTGCGGCTCCTCGGACCTTTACCTG 772  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141  
 DB 773 GTTAGCGACGCGCGAGCTCATTCCTGCGCGCGGAGGTGATACAGGGGTAGCTG 832  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161  
 DB 833 CTTTCGCCCGCGCCCATTTCTTACCTAAAGGCTCCTCGGGGGTCCGCTGTTGTGCCCC 892  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
 DB 893 GCGGACACGCGGTGGGCTATTACGGCCCGGGTGTACCCCGTGGAGTGGCCAGGCG 952  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 DB 953 GTGGACTTATCCCTGTGGAGAACCTAGACACACCATGAGATCC 997  
 RESULT 2  
 LOCUS ARL145264 1998 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 105 from patent US 6211338.  
 ACCESSION ARL145264  
 VERSION ARL145268.1 GI:15107135

ARL145264  
 VERSION ARL145264.1 GI:15107131  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 1998)  
 AUTHORS Malcolm B.A., Taremi S.Shane., Weber P.C. and Yao N.  
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide  
 JOURNAL Patent: US 6211338-A 105 03-APR-2001;  
 FEATURES  
 source Location/Qualifiers  
 1..1998 /organism="unknown"  
 BASE COUNT 411 a 595 c 569 g 423 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9,21e-67 Length: 1998  
 Score: 911.50 Matches: 170  
 Percent Similarity: 94.90% Conservative: 16  
 Best Local Similarity: 86.73% Mismatches: 7  
 Query Match: 88.32% Indels: 3  
 DB: 6 Gaps: 1  
 US-09-965-594-16 (1-197) x ARL145264 (1-1998)  
 QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTCTGTGTATTGTTGGTAGAATATTTTATCTGTTAGTGTAGTATACAGCGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACGCGGGGCTACTTGTGTTGCAAGAGACTAGCCTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTGGAGGGTGAGTGTTCAGGTGTCTCCACCGCAACAAATCTTCTCTGGGACCC 243  
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGGTCAACGCGGTGTGTGGACGCTTACCATGTGCTGCCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGCGCCATCCACCATGTACACTAATGTGACACGAGGACCTCGTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGCCCGCGGGCGGCTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCAGAGACATGCTGACGTCTATTCGGTGGCGCGCGGCGGACAGTAGGGGAGCCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161  
 DB 484 CPTCCCGGAGCGCTGTCTCTACTTTGAAGGGCTCTCCGGGTGTCCACTGCTGCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGGCTCTTCGGGCTGCCGTATGACCCCGGGGGTTCGGAAGGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTGTGCGGTAGTCCATGGAACTACTATGCGGTCTCG 651  
 RESULT 3  
 LOCUS ARL145268 1998 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 109 from patent US 6211338.  
 ACCESSION ARL145268  
 VERSION ARL145268.1 GI:15107135

## KEYWORDS

SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)  
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 109 03-APR-2001;  
FEATURES Location/Qualifiers  
1..1998

BASE COUNT 411 a 595 c 569 g 423 t  
ORIGIN /organism="unknown"

Alignment Scores:  
Pred. No.: 1.64e-66 Length: 1998  
Score: 908.50 Matches: 169  
Percent Similarity: 94.90% Conservative: 17  
Best Local Similarity: 86.22% Mismatches: 7  
Query Match: 88.03% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145268 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
DB 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTGTAGTATCATCGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
|||||  
DB 124 TCCCAACAGACGCGGGGCTACTTTGGTTGCAAGAAGATCAGCTTACAGCCCGGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
DB 184 ACCAGGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACAACTCTTCCGCGGACC 243  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
DB 244 TCGCTCAACGCGGCTGTGTTGACCGCTTACCATGCTGCTCAAGACCTTAGCGCGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACGAGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 364 GCGCCCGCGGGCGGCTTCTTGACACCATGCACCTGTGCACCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 424 CTCACGACACATGCTGACGTCTATTCGGTGGCGGGCGGCGACAGTAGGGGAGCCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
DB 484 CTCCTCCCGACGCTCTCTCTACTTGAAGGCTCTGCTGTGCTCCACTCTCTGCGCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
DB 544 TCGGGGACGCTGTGGGCTCTTCCGGCTCGGCTATGACACCGGGGGTTGCGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
DB 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 4  
LOCUS AR145262  
DEFINITION Sequence 103 from patent US 6211338.  
ACCESSION AR145262  
VERSION AR145262.1 GI:15107129  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

## ORGANISM

Unknown.  
Unclassified.

REFERENCE 1 (bases 1 to 1998)  
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 103 03-APR-2001;  
FEATURES Location/Qualifiers  
1..1998

BASE COUNT 410 a 596 c 568 g 424 t  
ORIGIN /organism="unknown"

Alignment Scores:  
Pred. No.: 1.98e-66 Length: 1998  
Score: 907.50 Matches: 170  
Percent Similarity: 94.39% Conservative: 15  
Best Local Similarity: 86.73% Mismatches: 8  
Query Match: 87.94% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145262 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
DB 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTGTAGTATCATCGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
|||||  
DB 124 TCCCAACAGACGCGGGGCTACTTTGGTTGCAAGAATCAGCTTACAGCCCGGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
DB 184 ACCAGGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACAACTCTTCCGCGGACC 243  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
DB 244 TCGCTCAACGCGGCTGTGTTGACCGCTTACCATGCTGCTCAAGACCTTAGCGCGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACGAGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 364 GCGCCCGCGGGCGGCTTCTTGACACCATGCACCTGTGCACCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 424 CTCACGACACATGCTGACGTCTATTCGGTGGCGGGCGGCGACAGTAGGGGAGCCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
DB 484 CTCCTCCCGACGCTCTCTCTACTTGAAGGCTCTGCTGTGCTCCACTCTCTGCGCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
DB 544 TCGGGGACGCTGTGGGCTCTTCCGGCTCGGCTATGACACCGGGGGTTGCGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
DB 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 5  
LOCUS AR145263  
DEFINITION Sequence 104 from patent US 6211338.  
ACCESSION AR145263  
VERSION AR145263.1 GI:15107130  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.



JOURNAL Patent: US 6211338-A 107 03-APR-2001:

FEATURES Location/Qualifiers

source 1..1998

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 3,53e-66 Length: 1998  
Score: 904.50 Matches: 169  
Percent Similarity: 94.39% Conservative: 16  
Best Local Similarity: 86.22% Mismatches: 8  
Query Match: 87.65% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x ARL45266 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTTCGTGTTATTTGTTAGTAATTTATTTATCTGTTAGTATCATCAGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
|||||  
Db 124 TCCCAACAGACGCGGGGCTACTTGGTGTGCAAGATCACTAGCCTTACAGCGCGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 184 AACCAAGTTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACAATCTCTCTGGCGACC 243  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 244 TCGCTCAACGCGGTGTGTGACCGTTTACCATGCTGTGCTGCTCAAGACCTTTAGCCGGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
Db 304 CCNAAAGGGCCCAATCACCAGATGTACACTAATGTGGACCAGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 GCGCCCCCGGGGCGGTCTCTTGCACACCATGTCACCTGTGCAGCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCCAGAGACATGCTGCTCCTACTTGAAGGCTCTGCTGGTCCACTGCTCTGCCCT 543  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCCTCCCCAGGCTGTCTCTACTTGAAGGCTCTGCTGGTCCACTGCTCTGCCCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
Db 544 TCGGGGCACGCTGTGGGCATCTCCGGCTCGCGCTATGCACCCGGGGGTTGCGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 8

ARL45267

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Sequence 108 from patent US 6211338.  
ARL45267.1 GI:15107134  
Unknown.  
Unclassified.  
1 (bases 1 to 1998)  
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
Patent: US 6211338-A 108 03-APR-2001;  
Location/Qualifiers

Sequence 102 from patent US 6211338.  
ARL45261.1 GI:15107128  
Unknown.  
Unclassified.  
1 (bases 1 to 1998)  
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
Patent: US 6211338-A 102 03-APR-2001;  
Location/Qualifiers

source

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 3,53e-66 Length: 1998  
Score: 904.50 Matches: 169  
Percent Similarity: 94.39% Conservative: 16  
Best Local Similarity: 86.22% Mismatches: 8  
Query Match: 87.65% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x ARL45267 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTTCGTGTTATTTGTTAGTAATTTATTTATCTGTTAGTATCATCAGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
|||||  
Db 124 TCCCAACAGACGCGGGGCTACTTGGTGTGCAAGATCACTAGCCTTACAGCGCGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 184 AACCAAGTTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACAATCTCTCTGGCGACC 243  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 244 TCGCTCAACGCGGTGTGTGACCGTTTACCATGCTGTGCTGCTCAAGACCTTTAGCCGGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
Db 304 CCNAAAGGGCCCAATCACCAGATGTACACTAATGTGGACCAGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 GCGCCCCCGGGGCGGTCTCTTGCACACCATGTCACCTGTGCAGCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCCAGAGACATGCTGCTCCTACTTGAAGGCTCTGCTGGTCCACTGCTCTGCCCT 543  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCCTCCCCAGGCTGTCTCTACTTGAAGGCTCTGCTGGTCCACTGCTCTGCCCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
Db 544 TCGGGGCACGCTGTGGGCATCTCCGGCTCGCGCTATGCACCCGGGGGTTGCGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 9

ARL45261

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Sequence 102 from patent US 6211338.  
ARL45261.1 GI:15107128  
Unknown.  
Unclassified.  
1 (bases 1 to 1998)  
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
Patent: US 6211338-A 102 03-APR-2001;  
Location/Qualifiers

Sequence 108 from patent US 6211338.  
ARL45267.1 GI:15107134  
Unknown.  
Unclassified.  
1 (bases 1 to 1998)  
Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
Patent: US 6211338-A 108 03-APR-2001;  
Location/Qualifiers



BASE COUNT 409 a 597 c 567 g 425 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4,28e-66 Length: 1998  
Score: 903.50 Matches: 170  
Percent Similarity: 93.88% Conservative: 14  
Best Local Similarity: 86.73% Mismatches: 9  
Query Match: 87.55% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGlyArgGileAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
DB 64 GGTTCTGTTGTTATTTGTTAGAAATTTATTTATCTGTTAGTAGTATCATCAGCGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
|||||  
DB 124 TCCCAACAGACGGGGGCTTACTTGGTTGTCATCATCACTAGCCTTACAGCGCGGGAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
DB 184 AACAGGTCAGGAGAGGTTTCAGTGGTTTCCACCGCAACAATCCTTCTCGCGGACC 243  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
DB 244 TGCCTCAACGGCGGTGTGTGGCGGTTTACATGTTGCTGGCTCAAAGACCTTAGCGCG 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
DB 304 CCAAGGGGCAATCACCAGATGTACATTAATGTGGACCGACCTCGTCGCGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 364 GCGCCCCCGGGCGGCTTCTTGCACACCATGCACCTGGCGACCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 424 GTCACGAGACATGTCAGTCATTCGGTGGCGCGGGCGGACAGTAGGGGAGCGCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
DB 484 CTCTCCCGAGCGCTGCTCCTACTTGAAGGGCTCTCGGGTGTCTCCTGCTCGCCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
DB 544 TCGGGGACCGTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
DB 604 GTGGACTTGTGGCCGTAGAGTCCATGGAACTACTATGCGGCTCGG 651

## RESULT 10

AR145269  
LOCUS AR145269 2016 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 110 from patent US 6211338.  
ACCESSION AR145269  
VERSION AR145269.1 GI:15107136  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

## REFERENCE

1 (bases 1 to 2016)  
AUTHORS Malcom,B.A., Tarem,S.,Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 110 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..2016  
/organism="unknown"

BASE COUNT 412 a 603 c 570 g 431 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4,32e-66 Length: 2016  
Score: 903.50 Matches: 170  
Percent Similarity: 93.88% Conservative: 14  
Best Local Similarity: 86.73% Mismatches: 9  
Query Match: 87.55% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgGileAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
DB 82 GGTTCTGTTGTTATTTGTTAGAAATTTATTTATCTGTTAGTAGTATCATCAGCGGCTAC 141  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
|||||  
DB 142 TCCCAACAGACGGGGGCTTACTTGGTTGTCATCATCACTAGCCTTACAGCGCGGGAAG 201  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
DB 202 AACAGGTCAGGAGAGGTTTCAGTGGTTTCCACCGCAACAATCCTTCTCGCGGACC 261  
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
DB 262 TCGCTCAACGGCGGTGTGTGGACGTTTACATGTTGCTGGCTCAAAGACCTTAGCGCG 321  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
DB 322 CCAAGGGGCAATCACCAGATGTACATTAATGTGGACCGACCTCGTCGCGCTGGCAG 381  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 382 GCGCCCCCGGGCGGCTTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 441  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 442 GTCACGACATGCTGACGCTCATTCGGTGGCGGGCGGACAGTAGGGGAGCGCTG 501  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
DB 502 CTCTCCCGAGCGCTGCTCCTACTTGAAGGGCTCTTCGGTGTCTCCTGCTCGCCT 561  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
DB 562 TCGGGGACCGTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 621  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
DB 622 GTGGACTTGTGGCCGTAGAGTCCATGGAACTACTATGCGGCTCTCG 669

## RESULT 11

AF268278  
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000  
DEFINITION Pestivirus type 1, complete genome.  
ACCESSION AF268278  
VERSION AF268278.1 GI:9049956  
KEYWORDS  
SOURCE  
ORGANISM

## Pestivirus type 1

Pestivirus type 1  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

## Pestivirus

1 (bases 1 to 12734)

## REFERENCE

AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassiliev,V.,  
Donis,R.O., Hong,Z. and Lau,J.Y.

## TITLE

Generation and characterization of a hepatitis C virus NS3  
protease-dependent bovine viral diarrhea virus

## JOURNAL

J. Virol. 74 (14), 6339-6347 (2000)

## MEDLINE

20323484

## PUBMED

10864644

## REFERENCE

2 (bases 1 to 12734)

## AUTHORS

Lai,V.C.H. and Hong,Z.

## TITLE

Direct Submission

JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

FEATURES Location/Qualifiers

source l..12734

1..12734

1..385

386..12508

/codon\_start=1

/product="polyprotein"

/protein\_id="AAF82566.1"

/db\_xref="gi:9049957"

/translation="MELTNEGSGSVIVIGRVLGSGSITACAQQRGLGCKITSL  
TGRKNQVEGVQIVSTATOTFLATCINGCVTIVHCAQTRITASPKGPVIOVITND  
ODLVGPAQGRSRLTCTCGSDPLVTRHANVIVRRRGDSGLSPRLSYLKG  
SSGGLPCAGHAGVGLRAAIVCTRGVAKAVDFIPVENLETTTSGSGADTEDVVCSSM  
SYSTKEGATKQKXPKRERKMIKPKSEKSKTKPPDATIVVEGVYVRRK  
GKTKSNTQDGLYHNKPKQESKKKALLAWAIIAIVLFVMTGENITOMNLQNG  
TEGIQAMFORGVNRSLHGIPEKICTGVPSHLATIDELTKIHGMDSASEKITVTCGR  
LORHEMKNKGCNWNTEPWILVMNRTOANLTEGQPPRECAVTCRYDRASDLNVTQA  
RDSPTPLTGCKKGNFSPAGILMRPCNFEIATASDVLFKHERILSMFQDITLYLVQGL  
TNSLEGARQGTAKLITLWLGKGLGKLENKSKTWFCAYAAAPYCDVDRKIGIYWT  
KNCIPACLPKTKIVGPKFOTNAEDGKILHEMGHLSHEVLLLSLVLSDFAPETASV  
MYLILHFSIPQSHVDVMDCKDTQNLTVELTITADVIPGSVMNLGKWICIRPNWMPYET  
TVTLAFEVSVQVWLRAALRDLRIWNAATTTAFCLVKIVRGQVOGILMLLIT  
CVQGHLDCKPEFSAIAKDERIGQGAELTTTWKEYSPGKLEDDTWIAWCEGDKLM  
VLCRTRETVYLAIIHTRALPTSVPFKLFDGRKQEDVEMDNFEGLCPCDAKPIV  
RGKPNITLLNGPAQWCPICWCTVYSCTSFNMOTLATTVVTVYRSGKPPHQQGJIT  
QKNIGEDJHNCILGGMWTCVPGDQLLYKGGIESCKWCGQFKESEGLPHYPKQCKL  
ENETGYRLVSTNSCNREGAVIPQGLTKCKTQVVIAMDTKLVDGPHMPKRPVEIIS  
EGPVEKTAFTNTYTKANKFYEPDSYFOQYMLKGVQYWFDPLEVDTHDRDFAES  
LYLVALLGGRVYLLWLVYMWLSEQALGIQGVGVVMGNLLTHNNIEVYTFLL  
LYLLRESVYKWLHLLHILVHPKISVIVILLMGVKAQSGGQYGLGKIDLCPT  
TVVLIVGLIILARQDPTVPLVITMAALRVTELTHOPQVDLAVAVMTITLLWYVTD  
YFRKKNLQCLISLVSGVFLNSLYLGRITEMPEVTPNWRPLTLLILYLSITIVTR  
WRVDVAGLLQCVPLILLVLTLMADFLLILIPYELVLYLKTVRTDIERSWLG  
IDYTDVSIYDDESGEYVLPFRQKAGNFSILLPLIKATLISCVSSKWLQYLSY  
LTLDPMYHRRVYIIESSGNTIISRLVAALIELNMEESKGLKFKYLLSGLRN  
L11KHKNVETVSNVSGEYVGMKIMTILKASTLSKRHCIICTCEGHEMKGTC  
PKCGRHGPICTGSLADEERHYKRFIREGNEGMSRCQGRHREPREDRPKSAR  
YCAECNRLHAEQGFPHAESHLGLITYPALADGKYDITENWAGCORVGLSPTRHY  
PCHISGSMFRQINGVQYTAGQLFRLNPLVATFKVAMLVNGBEIGLEHL  
GHILSPAVCKRIETHEKCHINILDKLTAFFGIMPRGTPPRAPVFTLSLLVRRLG  
TGWATTHOGGSSVDHYTAGDLVCDMSGRVYVCCSNRRLTDEYGVKTDSCQPD  
GARCVLNPEAVNTSGSGAVVHLQKTGDEFTCVTASGTPAFDLKLNKLSGLPIFE  
ASSGRVYVGVKNNEESKPTKIMSGIOTVSKNTADLTENVKKTISNMNRGDKOITLA  
TGAGRTTELPAVTEIGRHKRVLYLPLRAAESVQYMRLLKHPSTISNLRIGDKE  
GDMATGTTASVGYFCQMPKLRAMVVEYSYIFLDEHCAITPEQLAIIKIHRES  
IHVAMTATPAGSVTTTQCKHPIDEEFIAPENVKGGEDSGQFLDIAGLIPVDEMKNM  
LVFVPTRNMAVEVAKKAKAGYNSGYYSGEDPANLRVVTSSQSPYIVATNAIESGT  
LPDLTVITDGLCKEKRVRVSKIPFIVTGLKRMATVGEQAQRGRVGRVKGPRYR  
SOETATGSKVDYLLQAOYGIEDIQVNTKSFREMYDMSLYBEDSLITOLEILAN  
LLISDLPAVKNIMADHDPEIOLAYNSVEVQVPLFPKIRNGEVDTVYENSEFN  
ARKLEDVPIYIATEDLAVDLGLDMDPDGQOVVECTGKALQVNTGLSSAENAL  
VALGVYVQALSKRHVPITDITIEQORLEDTHLYAPNAIKDGTETELKELAS  
GVKEMIGASIDYAGGLFVKSQAEKTKTAPLFKENAAGKGVYVQKFDILSENKEE  
IIRGLMCTHTALYKIAARHETAFATLVKLWLAFCGESVDHVOQAADVILNXYV  
HNKPSFGDSETOQGRFVASFILSALATYTKYNTWYHNSKYVEPALAYLPYATSA  
LKMTFTPLRESVLYSTTYKTLIRKGSGLDGLTGISAAMEILISQNPVSQYISVM  
LGVGAIAAHNAIETSEOKRTLMVKFNFLDOATDELKVENPEKIMALFEAVOTI  
GNPLRLYHLQYNGKWEAKELMSERTAGRNLFILMEAFELLMGDSQGIKIRLSON  
YILDLYGLHUKYNGKMLVGNWAPAFSCDPTSDERTILPTDILURVETPCGY  
EMKAFKNVGGKUTKVEESGFFLCRRNPGRPVNYRYTYIDDNLEIKFPAKLBSQVE  
HYKGVATIDYSGKMLLATDKWEHGVITRLAKRYTGVGFNGAYLGDEPNHRLV  
EROCATITKNTQVFLKMKKCAFTYDLTISNLTIELVHRNNLEKEEIPITATVTL  
AYTEVNEGVTIKPVLGERVIPDPVVDINLOPEVOTISEVGIITIGRETITMTGTVP  
VLEKVEPDASDNQNSVKIGLDSGNYPGIQTHTLIEIHNRPDARPPFMIILGSRNIS  
NRAKTAPNKVITGNDPREIDLMAGRNLVVALRDVDPPELSEWDFKGTFLDREALE  
ALSQAPKNTQYTKKAVRNLIIEQKDVPEIPNFWASDDPVFLVAKLNDKYYLVGDGE  
VKDQAKAIGATQATRIIEKVGSRGYAMKLSWFLQASNKQSLTPLFEELLRCPAT  
KSNKHMASAYOLAQGNMEPLGCVHLGTIPARRVYIHPYAYLKLKDFIEEEKKPR  
VNDTVIREHNKILKIRIQGNLNTKMLNPGKLSQLDREGKRNINHQIGTIMSS

AGIRLEKLPVIRVADTDTKTFFHEAIKDKIKSENQNPENHKLLEIFHTIAQPLKHT  
YKNEARDELEAGINRKGAGFLEKKHVLSEKHLVEOLVRLDKAGKRTIKYYETAI  
PKEARVDDQMOAGDLVYKRPVIOYPEAKTRLAITKVMYNNVQOPVYPCYGR  
TPLNFIPKVRKENDSFNEPVAFDTRKANTQVTSKDLQILQELQIYYKKEHKKFI  
DTIDHMTFVPTADGEVYIRNGQSGQPDTSAGNSMLNVLTMWTFACFESTGVPYK  
SFNRHARIHVGDDGLITERGLGLKFKANGQILHEAGRPQKITEGEMKVAYRFED  
IEFCSHTPPVPRWSDNTSSHMAGRDATVILSKMATRLDSSGERGTTAYEKAFAFSFL  
MYSNPLVRRICLLVLSSQPETDPSKHATYYKGPDIKAYDKVIGRNLSELKRTGFEX  
LANLNLSTLGIWTIKTKTSKRIIDCVAIKGEENMLVNLNDRLLSSKTHLYIPDKGF  
TLOGKHYSQLOLRTETPNPMGVGTRYKLGPIVNLRLRLAKILLMTAVGVSS"

3'UTR  
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others  
ORIGIN  
12509..12734

Alignment Scores:  
Pred. No.: 3 99e-65 Length: 12734  
Score: 902.50 Matches: 177  
Percent Similarity: 93.33% Conservative: 5  
Best Local Similarity: 90.77% Mismatches: 10  
Query Match: 87.45% Indels: 3  
DB: 14 Gaps: 1

US-09-965-594-16 (1-197) x AP268278 (1-12734)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 413 GGTAGTGTGTTATTGTTGTTAGTAAATTTGTTTATCTGTTAGTGTAGTATCATCGCGTGC 472

Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
|||||  
Db 473 GCCACGACGACGAGAGGCTCTTAGGGGTGTAGATCACACCTGCTGACTGGCGCGGACAAA 532

Qy 42 AsnGlnValIleGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 533 AACCAAGTSGAGGTCAGCTCCAGATCGTGCACTGCTACCCAAACCTTCCTGCGCAACG 592

Qy 62 CysIleAsnGlyValCysThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 593 TGCATCAATGGGTATGCTGGACCTGTCTACCACTGCTGCTGCGGCGGACGAGGACCATCATCA 652

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
|||||  
Db 653 CCCAAGGTCCTGTATCATCAGATGATATACATGTGGACCAAGACCTTGTGGGCTGGCGCC 712

Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 713 GCTCCTCAAGGTTCCCGCTCATTTGACACCTGACACCTGCGGCTCTCGGACCTTTACCTG 772

Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 773 GTTACGAGGACGCCAACGTCATTCCTGCGCGCGGAGGTATAGCAGGGGTAGCGCTG 832

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 833 CTTTCGCGCGCGCCCATTTCTTACCTAAAAGGCTCCTCTGCGGGGTCCGCTGTTGTGCGCCC 892

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
Db 893 GCGGACACGCCCTGGGCGCTATTTCAGGCGCGCGGTGTGACCCGCTGAGTGGCCACGCGC 952

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 953 GTGACATTTATCCTGTGGAGAACCTAGACAGAACCCAGAGATCC 997

RESULT 12  
ARI45258  
LOCUS ARI45258  
DEFINITION Sequence 99 from patent US 6211338.  
ACCESSION ARI45258  
VERSION ARI45258.1 GI:15107125  
KEYWORDS Unknown.  
ORGANISM Unknown.

Unclassified.

## REFERENCE

1 (bases 1 to 651)

## AUTHORS

Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.

## TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

## JOURNAL

protease and NS4A cofactor peptide

## FEATURES

Patent: US 6211338-A 99 03-APR-2001;

## source

Location/Qualifiers

1..651

/organism="unknown"

## BASE COUNT

120 a 187 c 200 g 144 t

## ORIGIN

Alignment Scores:

Pred. No.: 1.83e-66 Length: 651

Score: 901.50 Matches: 168

Percent Similarity: 94.87% Conservative: 17

Best Local Similarity: 86.15% Mismatches: 7

Query Match: 87.35% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145258 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyGlyAsp-----ThrAlaTyr 21

Db 64 GGTCTCTGTTATTTGTTGGTAGAATATTTTATCTGGTAGTGGTAGTATCATCGGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGACGGGGCGCTACTTGGTTGCAAGACATAGCCCTTACAGCGCGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACAGATCCAGGAGAGGTTCCAGGTGGTTTCCACCGCAACACAATCCCTTCCTGGCGAOC 243

QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TGGCTCAACGGCGTGTGTGACCGTTTACCATGGTCTGGCTCAAAGACCTTAGCGCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Db 304 CCAAGGGGGCAATCACCCAGATGTACACTAATGTGGACACGACCTCGTCGGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 GCGCCCCCGGGCGGTCTCTGACACCAATGACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCCTCCCGCGGCGCTCTCTCTACTTGAAGGGCTCTCGGGTGGTCCACTGCTCGCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGATCTTCCGGCTGCGGCTGACCCCGGGGGTTCGAAGGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196

Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAATACTACTATGCGGTCT 648

## RESULT 13

## LOCUS

AR145252

Sequence 93 from patent US 6211338.

ACCESSION

AR145252

VERSION

AR145252.1 GI:15107119

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 651)

## AUTHORS

Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.

## TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

## JOURNAL

protease and NS4A cofactor peptide

## FEATURES

Patent: US 6211338-A 93 03-APR-2001;

## source

Location/Qualifiers

1..651

/organism="unknown"

## BASE COUNT

119 a 188 c 199 g 145 t

## ORIGIN

Alignment Scores:

Pred. No.: 2.21e-66 Length: 651

Score: 900.50 Matches: 169

Percent Similarity: 94.36% Conservative: 15

Best Local Similarity: 86.67% Mismatches: 8

Query Match: 87.26% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145252 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyGlyAsp-----ThrAlaTyr 21

Db 64 GGTCTCTGTTATTTGTTGGTAGAATATTTTATCTGGTAGTGGTAGTATCATCGGGCTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGACGGGGCGCTACTTGGTTGCAAGATCAGCCCTTACAGCGCGGACAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACAGATCCAGGAGAGGTTCCAGGTGGTTTCCACCGCAACACAATCCCTTCCTGGCGAOC 243

QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TGGCTCAACGGCGTGTGTGACCGTTTACCATGGTCTGGCTCAAAGACCTTAGCGCGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101

Db 304 CCAAGGGGGCAATCACCCAGATGTACACTAATGTGGACACGACCTCGTCGGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 GCGCCCCCGGGCGGTCTCTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCCTCCCGCGGCGCTCTCTCTACTTGAAGGGCTCTCGGGTGGTCCACTGCTCGCCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGATCTTCCGGCTGCGGCTGACCCCGGGGGTTCGAAGGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196

Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAATACTACTATGCGGTCT 648

## RESULT 14

## LOCUS

AR145253

Sequence 94 from patent US 6211338.

ACCESSION

AR145253

VERSION

AR145253.1 GI:15107120

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 651)

AUTHORS

Malcolm, B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide  
Patent: US 6211338-A 94 03-APR-2001;

JOURNAL  
FEATURES

source  
1..651  
Location/Qualifiers  
119 a 188 c 199 g 145 t

DRIGIN

Alignment Scores:

Pred. No.: 2,21e-66 Length: 651  
Score: 900.50 Matches: 169  
Percent Similarity: 94.36% Conservativeness: 15  
Best Local Similarity: 86.67% Mismatches: 8  
Query Match: 87.26% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145253 (1-651)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGGTAGAATTATTTATCTGTTAGTGTATCATCGGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGAGCGGGGCTACTTGGTTGCATCAGACTAGCCCTACAGCGCGGACAAG 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACATCTTCTCGGCGACC 243
|||||
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGTCACAGCGGTGTGTGGACCGTTTACCATGTTGCTGCTCAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGTCCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGAGACATGCTGACGTCATTCGGTGCAGCGGCGGCGAGACAGTAGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCCAGGCGCTCTCTTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyPheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
Db 604 GTGGACTTGTCCCGTAGAGTCCATCGAAACTACTATGCGGTCT 648
```

RESULT 15

AR145265  
LOCUS 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 106 from patent US 6211338.  
ACCESSION AR145265  
VERSION AR145265.1 GI:15107132

KEYWORDS

Unknown.  
Organism.

REFERENCE

1 (bases 1 to 1998)  
Malcolm B.A., Taremi, S. Shane., Weber, P.C. and Yao, N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 106 03-APR-2001;

JOURNAL

FEATURES

source  
1..1998  
Location/Qualifiers  
409 a 597 c 567 g 425 t

BASE COUNT  
ORIGIN

Alignment Scores:

Pred. No.: 7,61e-66 Length: 1998  
Score: 900.50 Matches: 169  
Percent Similarity: 93.88% Conservativeness: 15  
Best Local Similarity: 86.22% Mismatches: 9  
Query Match: 87.26% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-16 (1-197) x AR145265 (1-1998)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTTCGTGTTATTGTTGGTAGAATTATTTATCTGTTAGTGTATCATCGGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGAGCGGGGCTACTTGGTTGCATCATCTAGCCCTTACAGCGCGGACAAG 183
|||||
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACATCTTCTCGGCGACC 243
|||||
QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TCGTCACAGCGGTGTGTGGACCGTTTACCATGTTGCTGCTCAAGACCTTAGCGCGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCACGAGACATGCTGACGTCATTCGGTGCAGCGGCGGCGGCGAGACAGTAGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCCCAGGCGCTCTCTTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyPheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGCATCTTCCGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTGTCCCGCTAGAGTCCATCGAAACTACTATGCGGTCTCCG 651
```

Search completed: August 31, 2003, 00:46:14

Job time : 2570.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....YAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09965594/runat\_29082003.151918.28302/app\_query.fasta\_1.2872  
-DB=N\_Geneseq\_15Jun03 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGCN\_1\_1\_1412 -runat\_29082003.151918.28302 -NCP0=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1032	100.0	594	21	AA73331	Hepatitis C virus
2	1015	98.4	594	21	AA73330	Hepatitis C virus
3	1002	97.1	594	21	AA73332	Hepatitis C virus
4	990	95.9	594	21	AA73333	Hepatitis C virus
5	981	95.1	588	21	AA73329	Hepatitis C virus
6	980	95.0	594	21	AA73334	Hepatitis C virus
7	976	94.6	594	21	AA73335	Hepatitis C virus
8	942	91.3	588	21	AA73328	Hepatitis C virus
9	922.5	89.4	12734	24	ABA95615	Chimeric BVDV/HCV
10	911.5	88.3	1998	20	AA80355	HCV NS4A-NS3 compl
11	908.5	88.0	1998	20	AA80359	HCV NS4A-NS3 compl
12	907.5	87.9	1998	20	AA80353	HCV NS4A-NS3 compl
13	907.5	87.9	1998	20	AA80354	HCV NS4A-NS3 compl
14	904.5	87.6	612	25	ABX15706	Anti-viral synthet
15	904.5	87.6	651	20	AA80345	HCV NS4A-NS3 compl
16	904.5	87.6	1998	20	AA80357	HCV NS4A-NS3 compl
17	904.5	87.6	1998	20	AA80358	HCV NS4A-NS3 compl
18	903.5	87.5	1998	20	AA80352	HCV NS4A-NS3 compl
19	903.5	87.5	2013	20	AA80360	HCV NS4A-NS3 compl
20	901.5	87.4	651	20	AA80349	HCV NS4A-NS3 compl
21	900.5	87.3	651	20	AA80343	HCV NS4A-NS3 compl
22	900.5	87.3	651	20	AA80344	HCV NS4A-NS3 compl
23	900.5	87.3	1998	20	AA80356	HCV NS4A-NS3 compl
24	900.5	87.3	2016	20	AA80361	HCV NS4A-NS3 compl
25	900	87.2	648	20	AA80365	HCV NS4A-NS3 compl
26	898	87.0	648	20	AA80363	HCV NS4A-NS3 compl
27	897.5	87.0	650	20	AA80347	HCV NS4A-NS3 compl
28	897.5	87.0	651	20	AA80348	HCV NS4A-NS3 compl
29	897.5	87.0	651	20	AA80351	HCV NS4A-NS3 compl
30	896.5	86.9	651	20	AA80342	HCV NS4A-NS3 compl
31	894	86.6	648	20	AA80362	HCV NS4A-NS3 compl
32	893.5	86.6	650	20	AA80346	HCV NS4A-NS3 compl
33	893.5	86.6	651	20	AA80350	HCV NS4A-NS3 compl
34	891	86.3	8145	20	AA823259	Plasmid pET-BS(+)/HCV NS3 DNA. Hepa
35	889	86.1	1933	20	AA823258	Hepatitis C virus
36	888.5	86.1	9646	19	AAV59361	cDNA encoding hepa
37	888.5	86.1	9646	24	ABR87285	Hepatitis C virus
38	888.5	86.1	12980	19	AAV59364	Hepatitis C virus
39	888.5	86.1	12980	24	ABR87286	Hepatitis C virus
40	888.5	86.1	16622	21	AA236212	Nucleotide sequenc
41	884.5	85.7	5300	10	AA92097	Combined open read
42	884.5	85.7	5360	10	AA92097	Hepatitis C virus
43	884.5	85.7	6905	10	AA92103	Combined open read
44	884.5	85.7	7310	10	AA92106	Combined open read
45	884.5	85.7	7310	10	AA920336	Composite hepatiti

ALIGNMENTS

RESULT 1  
AA73331  
ID AAA73331 standard; DNA; 594 BP.  
XX  
XX  
AC AAA73331;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.  
XX  
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutain; ds.  
XX  
XX Hepatitis C virus.  
OS  
OS Synthetic.  
FH Key Location/Qualifiers



xx	Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;
Alignment Scores:	
Pred. No.:	3,33e-86 Length: 594
Score:	1015.00 Matches: 194
Percent Similarity:	98.48% Conservative: 0
Best Local Similarity:	98.48% Mismatches: 3
Query Match:	98.35% Indels: 0
DB:	21 Gaps: 0
US-09-965-594-16 (1-197) x AAA73330 (1-594)	
Qy	1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db	1 ATGMAAAAAAAAAAGGATCGGTGTTATCGTCGGCGGTATCAACCTGTCGGTGACACCGCT 60
Qy	21 TyrAlaGlnGlnThrArgGlyClnGluGlyCysGlnGluThrSerGlnThrGlyArgasp 40
Db	61 TAGCGTCAACAGACACGAGGTGAGGAGGTGGCCAGAAACCTCCAGACCGCGTGGTGC 120
Qy	41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db	121 AAAAACCAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCTGGCT 180
Qy	61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db	181 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACCGTGTGTTACCGTACCATCGCT 240
Qy	81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
Db	241 TCCCCGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGGC 300
Qy	101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db	301 CCGGCTCCGACGGTTCGGTTCCTGACCCGCGTGACCTGCGGTTCTCCGACCTGTAC 360
Qy	121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140
Db	361 CTGGTTACCCCGTCACGCTGAGCTTATCCGGTTCGTCGTCGTGCTGCTCCGCTGGTCC 420
Qy	141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db	421 CTGCTGTCCCGCGCTCCGATCTCTACCTGAAGAGGTTCTCCCGTGGTCCGCTGCTGC 480
Qy	161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
Db	481 CCGGCTGGTCACGCTGGTGTATCTTCGCTGCTGCTGTTTGCACCCGCGGTGGTGTGCTAA 540
Qy	181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db	541 GCTGTTGACTTCACTCCGCTTGAATCCCTCGAAACCCACCACCATGCGTTCCTCCGG 591
RESULT 3	
ID	AAA73332 standard; DNA; 594 BP.
XX	XX
AC	AAA73332;
XX	XX
DT	19-DEC-2000 (first entry)
XX	XX
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.
XX	XX
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX	XX
KW	liver failure; liver cancer; mutant; mutain; ds.
XX	XX
OS	Hepatitis C virus.
OS	Synthetic.
XX	XX
XX	XX
Key	Location/Qualifiers
FT	1..594
CDS	/tag- a
FT	/product- "NS4A-NS3 fusion protein #5"
FT	FT







PD 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX P-PSDB; AAB15220.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic

XX amino acid, useful for screening inhibitors that may treat hepatitis C

XX

XX Claim 26; Fig 12; 66pp; English.

XX

XX The present sequence is the coding sequence for a mutated version of a

XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A

XX protease enzymes. These proteins are both essential for the replication

XX of the virus, acting to cleave its replicative proteins from the

XX polyprotein produced from the HCV genome. Inhibitors of the two proteins

XX should be effective as antiviral treatments of HCV infection. This is

XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver

XX failure and liver cancer. The present invention concerns a number of NS3

XX mutants and NS3-NS4A fusion proteins which can be used to identify

XX inhibitors of this type, as well as enabling structural studies of the

XX protease and protease-inhibitor complexes. The protein produced from this

XX sequence contains the alpha-helix0-1 variant.

XX

SQ Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

Alignment Scores:

Prod. No.:	5,06e-83	Length:	588
Score:	981.00	Matches:	190
Percent Similarity:	96.95%	Conservative:	1
Best Local Similarity:	96.45%	Mismatches:	4
Query Match:	95.06%	Indels:	2
DB:	21	Gaps:	1

US-09-965-594-16 (1-197) x AAA73329 (1-588)

2y	1	MetLysLysLysGlySerValIleValGlyArgIleAsnLeuSerGlyAspThrAla	20
Db	1	ATGCAAAAAGAGTACGGTGTATCGTCGGCGGTATGCTACTGACGGT-----GCT	54
Qy	21	TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAsp	40
Db	55	TAGGCTCAGCACACTCGAGGTGAGGAGGTGGCCAGAAACCTCCCAAGACCGCTCGTGAC	114
Qy	41	LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla	60
Db	115	AAAAACCGAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCCCTGGCT	174
Qy	61	ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla	80
Db	175	ACCTGCATCAACGGGTGTTTGTCTGGACCGCTTTTACCACGGTGTGTGTACCCGTACCATCGCT	234
Qy	81	SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp	100
Db	235	TCCCCGAAAAGGTCGGGTATTCCAGATGTACACCAACCGTTGACAAAGACCTTGGTTGGTGG	294
Qy	101	GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr	120
Db	295	CCGGCTCCGACGGGTCCCGTTCCTGACCCCGTGCACCTCGCGTTCCTCCGACCTGTAC	354
2y	121	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer	140
Db	355	CTGGTACCGCTCACGCTGACGTTATCCGGTTCGTGCTCGTGGTGACTCCCGTGGTTC	414

Qy	141	LeuLeuSerProAqgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys	161		
Db	415	CTGCTGCCCGCGCCGATCTCTACTGAAAGGTCTCCCGGTGGTGGCTGCTGTC	474		
Qy	161	ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys	180		
Db	475	CCGGCTGGTCAGCGTGTGGTATCTTCGGTCTCTGTTGCACCGGTGTGTGCTAA	534		
Qy	181	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	197		
Db	535	GCTGTGTACTTCATCCGTTGAAATCCCTGGAAACACCATGGGTCCCGC	585		
RESULT 6					
AAAT3334					
ID	AAA73334 standard; DNA; 594 BP.				
AC	AAA73334;				
XX	19-DEC-2000 (first entry)				
DT	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.				
DE	Hepatitis; NS3 protease; viral replication; chronic liver disease;				
KW	liver failure; liver cancer; mutant; mutein; ds.				
KW	Hepatitis C virus.				
XX	Synthetic.				
OS					
XX	Key	Location/Qualifiers			
PH	CDS	1..594			
FT	/*tag= a				
FT	/product= "NS4A-NS3 fusion protein #7"				
XX	WO200040707-A1.				
PN	13-JUL-2000.				
XX	06-JAN-2000; 2000WO-US00345.				
XX	08-JAN-1999; 99US-0115271.				
XX	(BRIM ) BRISTOL-MYERS SQUIBB CO.				
PA	Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;				
PI	WPI: 2000-465976/40.				
DR	P-PSDB; AAB15225.				
DR	Modified hepatitis C virus (HCV) NS3 protease comprising at least 1				
XX	substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic				
PT	amino acid, useful for screening inhibitors that may treat hepatitis C				
PT	.				
XX	Claim 26; Fig 17: 66pp: English.				
PS	The present sequence is the coding sequence for a mutated version of a				
XX	fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A				
CC	protease enzymes. These proteins are both essential for the replication				
CC	of the virus, acting to cleave its replicative proteins from the				
CC	polyprotein produced from the HCV genome. Inhibitors of the two proteins				
CC	should be effective as antiviral treatments of HCV infection. This is				
CC	useful as HCV can lead to chronic liver disease such as cirrhosis, liver				
CC	failure and liver cancer. The present invention concerns a number of NS3				
CC	mutants and NS3-NS4A fusion proteins which can be used to identify				
CC	inhibitors of this type, as well as enabling structural studies of the				
CC	protease and protease-inhibitor complexes. The protein produced from this				
CC	sequence contains the alpha-helix0-7 variant.				
XX	Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;				
SQ					
Alignment Scores:					
Pred. No.:	6.36e-83	Length:	594		
Score:	980.00	Matches:	190		

Percent Similarity: 97.46% Conservative: 2  
 Best Local Similarity: 96.45% Mismatches: 5  
 Query Match: 94.96% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-16 (1-197) x AAA73334 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGAGATCCGTTGTTATCGTCGGCCGTATCAACCTGTCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 40  
 DB 61 TAGCGTCAGCAGACTCGAGGTCAGCAGGTCACCCAGAGACCTCCACACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAACCCAGGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGCCTTCTGGCT 180  
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTCATCAACGGTCTTCTGTGGACCGTTTACCACGGTGTGGTACCGTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysLeuValGlyTrp 100  
 DB 241 TCCCGGAAAGGTCCGGTTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CAGGCTCCGACAGGTCCCGTCCCGTACCCCGTGCACCTGGGGTCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCGTCGTGCTGCTGCTGCTGCTGCT 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGGTCGGATCTCTACCTGAAAGGTTCCTCCGGTGTGCTGCTGCTGCTG 480  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
 DB 481 CCGGCTGGTCACGCTGTGTGTATCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCTGTTGACTTATCCCGGTGAAATCCCTGGAACACCATCGTTCGCCG 591

RESULT 7

AAA73335  
 ID AAA73335 standard; DNA: 594 BP.

XX AC AAA73335;

XX DT 19-DEC-2000 (first entry)

XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.

XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 liver failure; liver cancer; mutant; mutein; ds.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT CDS

FT 1..594

FT /\*tag- a

FT /product- "NS4A-NS3 fusion protein #8"

XX PN W0200040707-A1.

XX PD 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US000345.

XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 DR WPI: 2000-465976/40.  
 DR P-PSDB; AAB15226.

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -

XX Disclosure; Fig 18; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0 wild-type sequence.

XX Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;

Alignment Scores:

Pred. No.: 1.51e-82 Length: 594  
 Score: 976.00 Matches: 189  
 Percent Similarity: 95.94% Conservative: 0  
 Best Local Similarity: 95.94% Mismatches: 8  
 Query Match: 94.57% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-16 (1-197) x AAA73335 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGAGATCCGTTGTTATCGTCGGCCGTATCAACCTGTCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgasp 40  
 DB 61 TAGCGTCAGCAGACTCGAGGTCGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAACCCAGGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTGGCT 180  
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTCATCAACGGTCTTCTGTGGACCGTTTACCACGGTGTGGTACCGTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysLeuValGlyTrp 100  
 DB 241 TCCCGGAAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CAGGCTCCGACAGGTTCCTGCTCCCGTACCCCGTGCACCTGGGGTCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGGTTACCGCTCACGCTGACGTTATCCCGGTTCGTCGTGCTGCTGCTGCTGCTGCTG 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGGTCGGATCTCTACCTGAAAGGTTCCTCCGGTGTGCTGCTGCTGCTG 480

QY 161 ProAlaGlyHisAlaValAlcIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
 DB 481 CCGGTGTCACGGCTGGGTATCTTCGCTGCTGCTTTCACACCGTGGTGTCTAA 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCG 591

# RESULT 8 AAA73328

ID AAA73328 standard; DNA; 588 BP.

AC AAA73328;

DT 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.

KW Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; ds.

OS Hepatitis C virus.

OS Synthetic.

Key Location/Qualifiers

FT 1..588

FT /tag= a

FT /product= "NS3-NS4A fusion protein"

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI; 2000-465976/40.

XX P-PSDB; AAB15212.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -

PS Disclosure; Fig 10; 66pp; English.

XX The present sequence is the coding sequence for a fusion protein created  
 CC using the hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.

XX Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

XX Alignment Scores:

Pred. No.: 2,29e-79 Length: 588  
 Score: 942.00 Matches: 185  
 Percent Similarity: 94.42% Conservative: 1  
 Best Local Similarity: 93.91% Mismatches: 9  
 Query Match: 91.28% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-16 (1-197) x AAA73328 (1-588)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGGTTCCGTTGTTATCGTCGCCCTATAGTACTGAACGGT-----GCT 54  
 QY 21 TTTATAGInGInThrArgGlyGluGluCysGInGluThrSerGInThrGlyArgASP 40  
 DB 55 TAGCTCAGCAGACTCGAGGCTGCTGGTTGCATCATCACCCTCCCTGACCGGTGCTGAC 114  
 QY 41 LysAsnGInValGluGluValGInIleValSerThrAlaThrGInThrPheLeuAla 60  
 DB 115 AAAACACAGGTTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCTCGCT 174  
 QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 175 ACCTGCATCAACGGTGTTCGCTGGACCGTTTACCACCGTGTGTTACCGGTACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGInMetTyrThrAsnValAspLysAspLeuValcIlyTrp 100  
 DB 235 TCCCGAAGGTCGCGTTATCCAGATGTACACCAAGCTTGACAAAGACCTGTTGGTTGG 294  
 QY 101 GlnAlaProGInGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 295 CCGGCTCCGACAGGTTCCGTTCCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 355 CTGGTTACCCGTCACGCTGACGTTATCCCGGTTCTGCTGCGTGAGACTCCCGGTGTTCC 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 415 CTGCTGTCGCCGCGTCGATCTCTACCTGAAAGGTTCCTCCGTTGTCGCTGCTGTCG 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
 DB 475 CCGGCTGGTCACGCTGTTGGTATCTCCGTTGCTGCTGTTGCACCGGTGTTGCTAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATGCGTTCCTCCG 585

## RESULT 9

ABA95615

ID ABA95615 standard; DNA; 12734 BP.

XX AC ABA95615;

XX AC ABA95615;

DT 21-MAR-2002 (first entry)

DE Chimeric BVDV/HCV NS3-wt sequence.

DE Pestivirus; Npro; protease; NS3; screening; ds.

XX Chimeric - Bovine viral diarrhea virus.

XX Chimeric - Hepatitis C virus.

XX US6326137-B1.

XX 04-DEC-2001.

XX 25-JUN-1999; 99US-0344456.

XX 25-JUN-1999; 99US-0344456.

XX (SCHE ) SCHERING CORP.

XX Hong 2, Lai VCH, Lau JYN;

XX WPI; 2002-121103/16.

XX Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3  
 PT protease gene, useful for in vivo screening of compounds which inhibit  
 XX HCV infection

XX Example 2: Columns 17-28; 20pp; English.

CC The present invention relates to a nucleic acid construct encoding a  
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
 CC each junction site recognised by the Npro protease is replaced with a  
 CC junction site recognised by the HCV NS3 protease. The construct is useful  
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
 CC protease, where screening may be in cell culture or in an animal model.  
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
 CC virus)/HCV NS3- $\psi$ , which was used to illustrate the present invention.

XX SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:  
 Pred. No.: 7.1e-76 Length: 12734  
 Score: 922.50 Matches: 180  
 Percent Similarity: 94.36% Conservativity: 4  
 Best Local Similarity: 92.31% Mismatches: 8  
 Query Match: 89.39% Indels: 3  
 DB: 24 Gaps: 1

US-09-965-594-16 (1-197) x ABA95615 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 413 GGTAGTGTGTATTTGTTGGTAGAATGTTTATCTGGTAGTGTAGTATCATCGCGGTAC 472  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 DB 473 GCCCAGCAGACGAGCGCTCTAGGTGTAGATCACCAGTCTGACTGCGCGGGACAAA 532  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 533 AACCAAGTGGAGGTGGTCCAGATGCTCAACTGCTACCCCAACCTTCTCGGCAAG 592  
 QY 62 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 593 TGCATCAATGGGTATGCTGGAGTGTCTACACCGGCGCGGACGAGGACCATCGCATCA 652  
 QY 82 ProGlyGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 653 CCCAAGGTCTCTCATCCAGATGTATACCAATCTGGACCAAGACCTTGTGGCTGGCC 712  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 713 GCTCTCAAGGTTCGGCTCATTTGACACCTTGACCTGCGGCTCTCGGACCTTTACCTG 772  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 773 GTTAGGAGCGACCGCGAGTCATTCCTGCTGCGCGGAGGTATAGCAGGGGTAGCCTG 832  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 833 CTTTGGCCCCCGCCCATTTCTACTATAAGGCTCTCTCGGGGGGTCTCGCTGTGTGCCCC 892  
 QY 162 AlaGlyHisAlaValAlaGlyIlePheAlaAlaValCysThrArgGlyValAlaLysAla 181  
 DB 893 CGGGGACACCGCGTGGGCTATTTCAGGGCGCGGTGTGCACCGGTGGAGTGGCAAGCG 952  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 DB 953 GTGGACTTTATCCCTGTGGAGAAGCTAGAGACAACCATGATGATCC 997

RESULT 10

AA80355

ID AAX80355 standard: cDNA; 1998 BP.

XX

AA80355;  
 XX 07-SEP-1999 (first entry)  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.  
 XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX WO9928482-A2.  
 XX 10-JUN-1999.  
 XX 24-NOV-1998; 98WO-US24528.  
 XX 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WPI; 1999-385385/32.  
 XX New hepatitis C virus covalent complexes  
 PT Disclosure; Page 166-169; 21pp; English.  
 PS The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.  
 XX SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:  
 Pred. No.: 7.6e-76 Length: 1998  
 Score: 911.50 Matches: 170  
 Percent Similarity: 94.90% Conservativity: 16  
 Best Local Similarity: 86.73% Mismatches: 7  
 Query Match: 88.32% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-16 (1-197) x AAX80355 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTCTGTGTATTTGTTGGTAGAATATTATTTCTGTGTAGTGTAGTATCATCGGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACGCGGGGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGGACAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTTCGAGGAGAGGTTTCAGTGGTTTCCACCGCAACAACTCTCTCTGGGACC 243  
 QY 62 CysIleAsnGlyValCysTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGGTCAACGGCGTGTGTGGACCTTTACCATGTGTGCTGCTCAAGACCTTAGCGGCG 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 DB 304 CCAAAAGGGCCCAATACCCAGATGACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGCCCGCCGGGGCGGCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGTCATTCCTGCTGAGGAGGCTTCCGGGTGGCCACTGCTGCCCC 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGGAGGCTGCTCTCTACTTGAAGGCTCTTCCGGGTGGCCACTGCTGCCCC 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCCGTATGCACCGGGGGTTCGCAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651  
 RESULT 11  
 AAX80359  
 ID AAX80359 standard; cDNA; 1998 BP.  
 XX  
 AC AAX80359;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN W09928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PE 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 XX  
 PR 28-NOV-1997; 97US-0067315.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX  
 DR WPI; 1999-385385/32.  
 XX  
 PT New hepatitis C virus covalent complexes  
 XX  
 PS Disclosure; Page 179-182; 211pp; English.  
 XX  
 CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

SO Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.45e-75 Length: 1998  
 Score: 908.50 Matches: 169  
 Percent Similarity: 94.90% Conservative: 17  
 Best Local Similarity: 86.22% Mismatches: 7  
 Query Match: 28.03% Indels: 3  
 DB: 20 Gaps: 1  
 US-09-965-594-16 (1-197) x AAX80359 (1-1998)  
 QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTTCGTGTTATTGTTGGTGGAGAAATTATTATCTGGTAGTGGTAGTATACCGGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCCAACACACGCGGGGCTACTTGGTTGCAAGAAGACTAGCTTACAGGCGGGCAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCCAGGTCGAGGGAGAGGTTCCAGGTGGTTCCACCGCAACACAAATCCTCTCGCGGACC 243  
 QY 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGGTCAACGCGCTGTGTGGACCGTTTACCATGCTGCTGGCTCAAGACCTTAGCCGCG 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 DB 304 CCNAAAGGGCCCAATACCCAGATGTACACTAATGTGGACCAAGACCTCTCGGTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGCCCGCCGGGGCGGCTTCCTTGACACCTGACCTGTGGCAGCTCAGACCTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGTCATTCGGGTGGCGGGCGGCGAGTAGGGGAGCCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGGAGGCTGCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTGCCGTATGCACCGGGGGTTCGCAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651  
 RESULT 12  
 AAX80353  
 ID AAX80353 standard; cDNA; 1998 BP.  
 XX  
 AC AAX80353;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN W09928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX

PF 24-NOV-1998; 98WO-US24528.  
XX  
XX 28-JUL-1998; 98US-0094331.  
PR 28-NOV-1997; 97US-0067315.  
XX  
XX PA (SCHE ) SCHERING CORP.  
XX  
XX MalcolM BA, Taremi SS, Weber PC, Yao N;  
XX WPI; 1999-385385/32.  
XX  
XX New hepatitis C virus covalent complexes  
XX  
XX Disclosure; Page 160-162; 21lpp; English.  
XX  
XX The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence encodes an example of the above complex. The covalent  
CC NS4A-NS3 complexes are useful for structural determination and  
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
CC They can also be used for detecting inhibitors of the protease activity,  
CC the helicase activity and the ATPase activity of NS3. The covalent  
CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
CC covalent protease-peptide complexes previously available.  
XX  
XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 1.8e-75 Length: 1998  
Score: 907.50 Matches: 170  
Percent Similarity: 94.39% Conservative: 15  
Best Local Similarity: 86.73% Mismatches: 8  
Query Match: 87.94% Indels: 3  
DB: 20 Gaps: 1  
US-09-965-594-16 (1-197) x AAX80353 (1-1998)  
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 64 GTTCTGTTGTTATTCTGCTAGCAATATTATTCTGTTAGTGGTAGTATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
Db 124 TCCCAACAGACGGCGGCTACTTGGTTGCAAGATCACTAGCCTTACAGCGCGGACAAAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCACCGCAACACATCTCTCTGGGACC 243  
Qy 62 CysIleAsnGlyValCysThrPrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 244 TGGGTCAACGGCGTGTGTGGACCGTTTACCATGGTGGCTCAAAGACCTTAGCGGCG 303  
Qy 82 ProLysGlyProValThrGlnMetThrThrAsnValAspLysLeuValGlyTrpGln 101  
Db 304 CCAAGGGGCAATACCCAGATGTACATAATGTGGACGAGACCTCTCGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 364 GCGCCCGCGGGCGGTTCTTGACACCATGACCTGTGGGAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 424 GTCAGAGACATCTGACGTCTATCCGGTCCGCGCGGCGGCGACAGTAGGGGACCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 484 CTCTCCCGGAGGCTGTCTCTCTACTTGAAGGGGCTCTTCGGGTGGTCCACTGCTGCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

Db 544 TCGGGGCACGCTGTGGGCATCTTCCGGGCTGCCGTATGCACCGGGGTTCGAAGGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
Db 604 GTGACTTGTGGCCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651  
RESULT 13  
AAX80354  
ID AAX80354 standard; cDNA; 1998 BP.  
XX  
XX AAX80354;  
XX  
XX 07-SEP-1999 (first entry)  
XX  
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.  
XX  
XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
XX  
XX Hepatitis C virus.  
OS Synthetic.  
XX  
XX WO9928482-A2.  
XX  
XX 10-JUN-1999.  
XX  
XX 24-NOV-1998; 98WO-US24528.  
XX  
XX 28-JUL-1998; 98US-0094331.  
PR 28-NOV-1997; 97US-0067315.  
XX  
XX (SCHE ) SCHERING CORP.  
XX  
XX MalcolM BA, Taremi SS, Weber PC, Yao N;  
XX WPI; 1999-385385/32.  
XX  
XX New hepatitis C virus covalent complexes  
XX  
XX Disclosure; Page 163-166; 21lpp; English.  
XX  
XX The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence encodes an example of the above complex. The covalent  
CC NS4A-NS3 complexes are useful for structural determination and  
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
CC They can also be used for detecting inhibitors of the protease activity,  
CC the helicase activity and the ATPase activity of NS3. The covalent  
CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
CC covalent protease-peptide complexes previously available.  
XX  
XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 1.8e-75 Length: 1998  
Score: 907.50 Matches: 170  
Percent Similarity: 94.39% Conservative: 15  
Best Local Similarity: 86.73% Mismatches: 8  
Query Match: 87.94% Indels: 3  
DB: 20 Gaps: 1  
US-09-965-594-16 (1-197) x AAX80354 (1-1998)  
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 64 GTTCTGTTGTTATTCTGTTAGCAATATTATTCTGTTAGTGGTAGTATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
Db 124 TCCCAACAGACGGCGGCTACTTGGTTGCAAGATCACTAGCCTTACAGCGCGGACAAAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 184 AACAGGTGAGGAGAGGTTCAAGTGGTTTCACCGCAACACATCTCTCTGGGACC 243  
Qy 62 CysIleAsnGlyValCysThrPrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 244 TGGGTCAACGGCGTGTGTGGACCGTTTACCATGGTGGCTCAAAGACCTTAGCGGCG 303  
Qy 82 ProLysGlyProValThrGlnMetThrThrAsnValAspLysLeuValGlyTrpGln 101  
Db 304 CCAAGGGGCAATACCCAGATGTACATAATGTGGACGAGACCTCTCGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 364 GCGCCCGCGGGCGGTTCTTGACACCATGACCTGTGGGAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 424 GTCAGAGACATCTGACGTCTATCCGGTCCGCGCGGCGGCGACAGTAGGGGACCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 484 CTCTCCCGGAGGCTGTCTCTCTACTTGAAGGGGCTCTTCGGGTGGTCCACTGCTGCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

```

Db 124 TCCCAACAGACGGGGCCCTACTTGGTGCATCAAGACTAGCCTTACAGCCGGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGGTTTCCACCGCAACAACTCTCTCTGGCGACC 243
Qy 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TGCCTCAACCGCGGTGGTGGACCGGTTTACCATGGTGGCTCAAGACCTTACCCCGGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGAGGCGCAATCAACCCAGATGTACACTAATGTGACAGACCTCTCTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCGCCGGCGGCTTCTTTCACACCATGACCTGTGGCAGCTTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTCATTCGGGTGCGCGCGGGGCGACAGTAGGGGAGCGCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCTCCCGCCAGCGCTGCTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACGCTGTGGGCATCTCCGGGCTGCGGTATGACACCGGGGGTTCGGAAGCGG 603
Qy 182 ValAspPheLeuProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACTTTGTGCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

RESULT 14
ABX15706
ID ABX15706 standard; DNA; 612 BP.
XX
AC ABX15706;
XX
DT 28-MAR-2003 (first entry)
XX
DE Anti-viral synthetic prototoxophore associated DNA sequence.
XX
KW Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;
KW virus; infection; antitumour; toxophore; human immunodeficiency virus;
KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.
XX
OS Unidentified.
XX
PN WO200287500-A2.
XX
PD 07-NOV-2002.
XX
PF 26-APR-2002; 2002WO-USI3223.
XX
PR 27-APR-2001; 2001US-286893P.
XX
PA (NEWB-) NEWBIOTICS INC.
XX
PI Cathers BE, Neuteboom STC, Shepard HM;
XX
PP WPI; 2003-167102/16.
XX
PT Novel synthetic viral prototoxophore for treating viral infections, has
PT toxin moiety incorporated into substrate domain specific for viral
PT enzyme, bound and modified by viral enzyme to get converted into
PT toxophore -
XX
PS Example 1; Page 62; 66pp; English.
XX
CC This invention relates to a novel synthetic viral prototoxophore

```

```

CC comprising a toxin moiety operatively incorporated into a substrate
CC domain specific for a viral enzyme. This prototoxophore may be bound
CC and modified by the viral enzyme thus converting it to a toxophore.
CC Also disclosed in the invention is a method for enhancing the anti-viral
CC effect of an antiviral agent, this method comprises contacting a cell,
CC infected with a virus or is susceptible to infection, with a
CC prototoxophore. The invention further comprises an assay to identify
CC anti-viral agents, comprising contacting an infected cell with a
CC candidate agent and comparing the ability of the agent to inhibit the
CC growth or infectivity of the virus in the cell. The prototoxophores
CC of the invention may have virucide or antitumour activity. The
CC prototoxophores of the invention may be useful for reducing or
CC inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte,
CC nerve cell, connective tissue cell, muscle cell or hepatocyte) which is
CC infected with a virus or is susceptible to infection with a virus, with
CC an effective amount of the prototoxophore. The cells are cell lines
CC adapted to long term continuous culture or isolated from a subject.
CC The prototoxophore is also useful for ameliorating the severity of a
CC viral infection in a subject, where the virus is selected from human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and
CC hepatitis virus, by administering an effective amount of the
CC prototoxophore to the subject. The prototoxophores of the invention are
CC also useful for treating tumours. The present sequence represents an
CC antiviral prototoxophore associated DNA sequence, this sequence is
CC described as a recombinant NS3/NS4 fusion protein in example 1 of
CC the invention although it is clearly not a protein sequence.
XX
SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 7.89e-76 Length: 612
Score: 904.50 Matches: 178
Percent Similarity: 92.82% Conservative: 3
Best Local Similarity: 91.28% Mismatches: 11
Query Match: 87.65% Indels: 3
DB: 25 Gaps: 1

```

US-09-965-594-16 (1-197) x ABX15706 (1-612)

```

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GGTAGTGTGGTGCATTTGGGTAGGATCATTTTGCCTAGTGGTAGTATCATCGCGGTAC 78
Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
Db 79 GCCCAGCAGACAAGGGGCTTCCTAGGTGCATATCACCAGGCTAACTGCCCGGGACAAA 138
Qy 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTGGAGGTGAGGTCCAGATTGTCTACCTGCTGCCCAACCTTCTCTGGCAAG 198
Qy 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTGCTGGACTGTCTACCGGGCGCGAAGCAGGACCATCGCGTCA 258
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 259 CCCAAGGCTCTGTCAATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGGCC 318
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGCAAGGTACCCGCTCATTTGACACCTCTGCGGCTCTCTGGACCTTTTACCTG 378
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGGACCGCGATGTCTTCCTGCGCGCGGGGGTGATAGCAGGCGGACGCTG 438
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGTGCGCCCGCGCCCATTTCTTGAAGGGCTCTCTCGGGGGGGTCCGCTGTGTGCCCC 498
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181

```

Db 499 CGGGGCGACGCCGTGGGCAATATTAGGGCGCGGTGTGCACCGTGGAGTGGCTAAGGCG 558

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 559 GTGGAGTTATCCCTGTGGAGAACCTAGACACAACCATGAGGTCC 603

RESULT 15  
AAx80345  
ID AAX80345 standard; cDNA: 651 BP.  
XX AC AAX80345;  
XX DT 07-SEP-1999 (first entry)  
XX DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.  
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
XX OS Hepatitis C virus.  
OS Synthetic.  
XX PN WO928482-A2.  
XX PD 10-JUN-1999.  
XX PF 24-NOV-1998; 98WO-US24528.  
XX PR 28-JUL-1998; 98US-0094331.  
XX PR 28-NOV-1997; 97US-0067315.  
XX PA (SCHE ) SCHERING CORP.  
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
XX DR WPI: 1999-385385/32.  
XX PT New hepatitis C virus covalent complexes  
XX PS Disclosure; Page 147-148; 21pp; English.  
XX CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence encodes an example of the above complex. The covalent  
CC NS4A-NS3 complexes are useful for structural determination and  
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
CC They can also be used for detecting inhibitors of the protease activity,  
CC the helicase activity and the ATPase activity of NS3. The covalent  
CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
CC covalent protease-peptide complexes previously available.  
XX SQ Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;

Alignment Scores:  
Pred. No.: 8,52e-76 Length: 651  
Score: 904.50 Matches: 169  
Percent Similarity: 94.87% Conservative: 16  
Best Local Similarity: 86.67% Mismatches: 7  
Query Match: 87.65% Indels: 3  
DB: 20 Gaps: 1

US-09-965-594-16 (1-197) x AAX80345 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTTCTGTTATTCTGGTAGAATATTTTCTGGTAGTGGTAGTATCAGGCCCTAC 123  
|||||

Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
:::|||||  
:::|||||

Db 124 TCCNACAGACGGCGGGGCTACTTGGTTGCAGAGACTAGCCTTACAGGCGGGGACAAG 183

Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 184 AACCAAGGTTCGAGGAGAGGTTACAGTGGTTTCCACCGCAACAATCCTTCCTGGCGACC 243

Qy 62 CysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 244 TGCGTCAACGGCGGTGTGTGGACGTTTACCATGGTGTGGCTCAAAGACCTTAGCCGCG 303

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
Db 304 CCAAAGGGGCCAATCACCCAGATGTACACTAATGTGCACGAGACCTCGTCGCGCTGCCAG 363

Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 GCGGCCCCCGGGGCGCGTCTTGCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423

Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCAGGACACATGCTGAGTCAATTCGCGCGCGCGGCGGCGACAGTAGGGGAGCGTG 483

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCCTCCCCAGGCGCTGCTCCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
|||||  
Db 544 TCGGGGCGACGCTGTGGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTTCGCAAGGCG 603

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGCCAAACTACTATGCGGTCT 648

Search completed: August 30, 2003, 19:48:03  
Job time : 188.939 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:26:03 ; Search time 176.482 Seconds  
(without alignments)  
2560.981 Million cell updates/sec

Title: US-09-965-594-16

Perfect score: 1032

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151920\_28367/app.query.fasta\_1.2872  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BL0sum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=pt0 -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09965594 @cgn\_1.1.864 @runat\_29082003\_151920\_28367  
-NCPU=6 -ICPU=3 -NO\_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1032	100.0	594	10	US-09-965-594-17 Sequence 17, Appl

2	1015	98.4	594	10	US-09-965-594-15	Sequence 15, Appl
3	1002	97.1	594	10	US-09-965-594-19	Sequence 19, Appl
4	990	95.9	594	10	US-09-965-594-21	Sequence 21, Appl
5	981	95.1	588	10	US-09-965-594-13	Sequence 13, Appl
6	980	95.0	594	10	US-09-965-594-23	Sequence 23, Appl
7	976	94.6	594	10	US-09-965-594-25	Sequence 25, Appl
8	942	91.3	588	10	US-09-965-594-4	Sequence 4, Appl
9	904.5	87.6	612	14	US-10-133-133A-6	Sequence 6, Appl
10	888.5	86.1	9846	9	US-09-742-653-3	Sequence 3, Appl
11	888.5	86.1	9846	10	US-09-238-076-1	Sequence 1, Appl
12	888.5	86.1	9846	11	US-09-995-937-1	Sequence 1, Appl
13	888.5	86.1	9846	11	US-09-917-563-1	Sequence 1, Appl
14	888.5	86.1	12980	10	US-09-238-076-5	Sequence 5, Appl
15	888.5	86.1	12980	11	US-09-995-937-5	Sequence 5, Appl
16	888.5	86.1	12980	11	US-09-917-563-5	Sequence 5, Appl
17	884.5	85.7	9379	9	US-09-916-359-1	Sequence 1, Appl
18	884.5	85.7	9416	10	US-09-238-076-19	Sequence 19, Appl
19	884.5	85.7	9416	11	US-09-995-937-19	Sequence 19, Appl
20	884.5	85.7	9416	11	US-09-917-563-19	Sequence 19, Appl
21	882	85.5	549	10	US-09-965-594-2	Sequence 2, Appl
22	882	85.5	2058	10	US-09-881-654-1	Sequence 1, Appl
23	882	85.5	2058	10	US-09-881-239-2	Sequence 2, Appl
24	881.5	85.4	836	10	US-09-921-337-120	Sequence 120, App
25	881.5	85.4	10803	10	US-09-747-419-17	Sequence 17, Appl
26	881.5	85.4	10803	14	US-10-259-275-17	Sequence 17, Appl
27	878.5	85.1	9416	10	US-09-929-955-13	Sequence 13, Appl
28	878.5	85.1	9416	13	US-10-104-966-13	Sequence 13, Appl
29	878	85.1	2061	10	US-09-929-955-16	Sequence 16, Appl
30	866.5	84.0	13910	11	US-09-919-901-1	Sequence 1, Appl
31	863.5	83.7	13910	11	US-09-919-901-8	Sequence 8, Appl
32	863.5	83.7	13910	11	US-09-919-901-15	Sequence 15, Appl
33	863	83.6	2064	11	US-09-884-456-69	Sequence 69, Appl
34	863	83.6	2523	11	US-09-884-456-85	Sequence 85, Appl
35	860.5	83.4	2073	14	US-10-133-133A-5	Sequence 5, Appl
36	860	83.3	6189	14	US-10-259-275-41	Sequence 41, Appl
37	860	83.3	7992	13	US-10-005-469-1	Sequence 1, Appl
38	860	83.3	7992	13	US-10-005-469-2	Sequence 2, Appl
39	860	83.3	7992	13	US-10-005-469-4	Sequence 4, Appl
40	860	83.3	7992	13	US-10-005-469-6	Sequence 6, Appl
41	860	83.3	8638	12	US-10-309-561-24	Sequence 24, Appl
42	860	83.3	8638	13	US-10-029-907-24	Sequence 24, Appl
43	860	83.3	8639	12	US-10-309-561-1	Sequence 1, Appl
44	860	83.3	8639	13	US-10-029-907-1	Sequence 1, Appl
45	860	83.3	8642	12	US-10-309-561-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-965-594-17  
; Sequence 17, Application US/09965594  
; Patent No. US20020106642A1  
; GENERAL INFORMATION:  
; APPLICANT: Wittekind, Michael  
; APPLICANT: Weinheimer, Steven  
; APPLICANT: Zhang, Yaqun  
; APPLICANT: Goldfarb, Valentina  
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for Facilitating Inhibitor Screening and Structural Studies  
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies  
; FILE REFERENCE: DB17Sequences  
; CURRENT APPLICATION NUMBER: US/09/965,594  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/115,271  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-17



; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-19

Alignment Scores:  
Pred. No.: 5,45e-107 Length: 594  
Score: 1002.00 Matches: 194  
Percent Similarity: 98.48% Conservatives: 0  
Best Local Similarity: 98.48% Mismatches: 3  
Query Match: 97.09% Indels: 0  
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-19 (1-594)

```
QY 1 MetLysLysLysGlySerValValleValGlyArgIleAsnLeuSerGlyAspThrAla 20
   |||
Db 1 ATGAAAAAAGGATCGTGTATCGTCGGCGTATCAACCTGTCGGGTGACACCGCT 60
   |||
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
   |||
Db 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTGCAGAAACCTCCAGACGGTCTGTGAC 120
   |||
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
   |||
Db 121 AAAAACCAAGTTGAAGTGAAGTTGAGATCGTTTCCACCGCTACCCAGACCTTCTCGCT 180
   |||
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
   |||
Db 181 ACCTCCATCAACCGGTGTTCTGTGACCGCTTACCACCGGTGCTGATCCGTACCATCGCT 240
   |||
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
   |||
Db 241 TCCCGGAAAGTCCCGTTACCCAGATGTACACCAAGCTTGACAAAGACCTGGTGGTTGG 300
   |||
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
   |||
Db 301 CAGGCTCCGACGGGTCCCGTCCCTGACCGCGTACCGCTGCTGCTGCTGCTGCTGCTG 360
   |||
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
   |||
Db 361 CTGGTTACCGGTACCGGTGACGTATATCCCGGTTCGTGCTGCTGCTGCTGCTGCTGCT 420
   |||
QY 141 LeuLeuSerProArgProLysSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
   |||
Db 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCCTCCGTTGCTGCTGCTGCTGCT 480
   |||
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
   |||
Db 481 CCGGCTGGTACAGCTGTGTTGATCTTCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
   |||
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
   |||
Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACCAACCATGCGTTCCTCCCG 591
```

## RESULT 4

US-09-965-594-21

; Sequence 21, Application US/09965594  
; Patent No. US20020106642A1  
; GENERAL INFORMATION:  
; APPLICANT: Wittekind, Michael  
; APPLICANT: Weinheimer, Steven  
; APPLICANT: Zhang, Yaqu  
; APPLICANT: Goldfarb, Valentina  
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for  
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies  
; TITLE OF INVENTION: of Protease:Inhibitor Complexes  
; FILE REFERENCE: DB17Sequences  
; CURRENT APPLICATION NUMBER: US/09/965, 594  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 21  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-21

Alignment Scores:  
Pred. No.: 1.34e-105 Length: 594  
Score: 990.00 Matches: 191  
Percent Similarity: 97.97% Conservatives: 2  
Best Local Similarity: 96.95% Mismatches: 4  
Query Match: 95.93% Indels: 0  
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-21 (1-594)

```
QY 1 MetLysLysLysGlySerValValleValGlyArgIleAsnLeuSerGlyAspThrAla 20
   |||
Db 1 ATGAAAAAAGGATCGTGTATCGTCGGCGTATCAACCTGTCGGGTGACACCGCT 60
   |||
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
   |||
Db 61 TACGCTCAGCAGACTCGAGGTGAGAGGGTGCAGAAAGCTCCACACCGTCTCCCTGGCT 120
   |||
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
   |||
Db 121 AAAAACCAAGTTGAAGTGAAGTTGAGATCGTTTCCACCGCTACCCAGACCTTCTCGCT 180
   |||
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
   |||
Db 181 ACCTCCATCAACCGGTGTTCTGTGACCGCTTACCACCGGTGCTGATCCGTACCATCGCT 240
   |||
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
   |||
Db 241 TCCCGGAAAGTCCCGTTACCCAGATGTACACCAAGCTTGACAAAGACCTGGTGGTTGG 300
   |||
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
   |||
Db 301 CAGGCTCCGACGGGTTCGGTTCCTGACCGCGTACCGCTGCTGCTGCTGCTGCTGCTG 360
   |||
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
   |||
Db 361 CTGGTTACCGGTACCGGTGACGTATATCCCGGTTCGTGCTGCTGCTGCTGCTGCTGCT 420
   |||
QY 141 LeuLeuSerProArgProLysSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
   |||
Db 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCCTCCGTTGCTGCTGCTGCTGCT 480
   |||
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
   |||
Db 481 CCGGCTGGTACAGCTGTGTTGATCTTCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
   |||
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
   |||
Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACCAACCATGCGTTCCTCCCG 591
```

## RESULT 5

US-09-965-594-13

; Sequence 13, Application US/09965594  
; Patent No. US20020106642A1  
; GENERAL INFORMATION:  
; APPLICANT: Wittekind, Michael  
; APPLICANT: Weinheimer, Steven  
; APPLICANT: Zhang, Yaqu  
; APPLICANT: Goldfarb, Valentina  
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for  
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies  
; TITLE OF INVENTION: of Protease:Inhibitor Complexes  
; FILE REFERENCE: DB17Sequences  
; CURRENT APPLICATION NUMBER: US/09/965, 594  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR FILING DATE: 1999-01-08  
; PRIOR FILING DATE: 1999-01-08

```

; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-13

Alignment Scores:
Pred. No.: 1.46e-104 Length: 588
Score: 981.00 Matches: 190
Percent Similarity: 96.95% Conservative: 1
Best Local Similarity: 96.45% Mismatches: 4
Query Match: 95.08% Indels: 2
DB: 1 Gaps: 1

US-09-965-594-16 (1-197) x US-09-965-594-13 (1-588)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGATCGGTGTTATCGTCGGCGGTATAGTACTGACCGT-----GCT 54
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 55 TACGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAGACCTCCAGACCGGTGCTGAC 114
QY 41 LysAsnGlnValGluGlyCyluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 115 AAAAACCCAGGTGGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACTTCCTGGCT 174
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 175 ACCTGCATCAACGGTGTTCGTGGACCGTTTACCACCGTGTGTTACCGCTACCATCGCT 234
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 235 TCCCGGAAAGTCCGGTTATCCAGATGTACCAACCGTTGACAAAGACCTGGTTGGTTGG 294
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 295 CCGGCTCCGACGGTTCCTCCGTCACCGCGTCCACCTCGCGTTCCTCCGACCTGTAC 354
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 355 CTGGTTACCGGTACGCTGACGTTATCCCGGTTCGTCGTGCTGACCTCCCGGTGTC 414
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 415 CTGCTGTCCCGGTCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 474
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 475 CCGGCTGGTACGCTGTTGTTATCTCCGCTGCTGTTGACCCCGGTGGTGGTGTAA 534
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 535 GCTGTTGACTTCATCCCGGTGAAATCCCTGGAAACCAACCATCGCTTCCCG 585

RESULT 6
US-09-965-594-23
; Sequence 23, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/1115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-23

Alignment Scores:
Pred. No.: 1.93e-104 Length: 594
Score: 980.00 Matches: 190
Percent Similarity: 97.46% Conservative: 2
Best Local Similarity: 96.45% Mismatches: 5
Query Match: 94.96% Indels: 0
DB: 1 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-23 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGATCGGTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTGACGAGGTACCCAGAAGACCTCCACACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTGGAAGGTGAAGTTCAGATCGTTTCCACCGCTTCCAGACCTTCCCTGGCT 180
QY 61 ThrCysIleAsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGTGTTCGTGGACCGTTTACCACCGTGTGTTACCGCTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGGAAAGTCCGGTTATCCAGATGTACCAACCGTTGACAAAGACCTGGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGACGGTTCCTCCGTCACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGGTACGCTGACGTTATCCCGTTCCGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGGTCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGCTGCT 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTACGCTGTTGTTATCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTGAAATCCCTGGAAACCAACCATCGCTTCCCG 591

RESULT 7
US-09-965-594-25
; Sequence 25, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease: Inhibitor Complexes
; FILE REFERENCE: DB17sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; CURRENT FILING DATE: 2001-09-27

```

; CURRENT APPLICATION NUMBER: US/09/965,594  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/115,271  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-25

Alignment Scores:  
Pred. No.: 5,62e-104 Length: 594  
Score: 976.00 Matches: 189  
Percent Similarity: 95.94% Conservative: 0  
Best Local Similarity: 95.94% Mismatches: 8  
Query Match: 94.57% Indels: 0  
DB: 10 Gaps: 0

US-09-965-594-16 (1-197) x US-09-965-594-25 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
DB 1 ATCAAAAAAAGGATCGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGCT 60  
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40  
DB 61 TACGCTCAGCAGACTCGAGGTCCTGGTTCATCATCCTCCCTGACCGGTGTCGTGAC 120  
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
DB 121 AAAAACCAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTCTCTGGCT 180  
QY 61 ThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAla 80  
DB 181 ACCTGTCATCAACCGGTGTTGCTGACCGCTTTACACCGGTGCTGTCACCGTACCATCGCT 240  
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
DB 241 TCCCGGAAGGTCGGTATCCAGATGTACACCAAGCTTGACAAGACCTGGTGGTGG 300  
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
DB 301 CCGGCTCGCAGGGTTCGGTTCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 360  
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
DB 361 CTGGTTACCGGTACCGCTGACGTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160  
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTACCTGAAGGTTCTCCCGGTGGTGGTGGTGG 480  
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
DB 481 CCGGCTGGTCAGCGTGGTGGTATCTTCGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATCGCTTCCCGC 591

## RESULT 8

US-09-965-594-4  
; Sequence 4, Application US/09965594  
; Patent No. US20020106642A1  
; GENERAL INFORMATION:  
; APPLICANT: Wittekind, Michael  
; APPLICANT: Weinheimer, Steven  
; APPLICANT: Zhang, Yaqun  
; APPLICANT: Goldfarb, Valentina  
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for  
; Facilitating Inhibitor Screening and Structural Studies

; TITLE OF INVENTION: of Protease Inhibitor Complexes  
; FILE REFERENCE: DB17Sequences  
; CURRENT APPLICATION NUMBER: US/09/965,594  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/115,271  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 588  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-4

Alignment Scores:  
Pred. No.: 4,84e-100 Length: 588  
Score: 942.00 Matches: 185  
Percent Similarity: 94.42% Conservative: 1  
Best Local Similarity: 93.91% Mismatches: 9  
Query Match: 91.28% Indels: 2  
DB: 10 Gaps: 1

US-09-965-594-16 (1-197) x US-09-965-594-4 (1-588)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
DB 1 ATCAAAAAAAGGATCGTTGTTATCGTCGGCGGTATAGTACTGAACGGT-----GCT 54  
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40  
DB 55 TACGCTCAGCAGACTCGAGGTCCTGGTTCATCATCCTCCCTGACCGGTGTCGTGAC 114  
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
DB 115 AAAAACCAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTCTCTGGCT 174  
QY 61 ThrCysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAla 80  
DB 175 ACCTGTCATCAACCGGTGTTGCTGGACCGTTTACCACGGTGTGTTACCGTACCATCGCT 234  
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
DB 235 TCCCGGAAGGTCGGTATCCAGATGTACACCAAGCTTGACAAGACCTGGTGGTGG 294  
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
DB 295 CCGGCTCGCAGGGTTCGGTTCCTGACCGCGTGCACCTCGGTTCTCCGACCTGTAC 354  
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
DB 355 CTGGTTACCGGTACCGCTGACGTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 414  
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160  
DB 415 CTGCTGTCCCGCGTCCGATCTCTACCTACCTGAAGGTTCTCCCGGTGGTGGTGGTGG 474  
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLys 180  
DB 475 CCGGCTGGTCAGCGTGGTGGTATCTTCGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 534  
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
DB 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAAACCAACCATCGCTTCCCGC 585

## RESULT 9

US-10-133-133A-6  
; Sequence 6, Application US/10133133A  
; Publication No. US20030114385A1  
; GENERAL INFORMATION:  
; APPLICANT: CATHERS, Brian  
; APPLICANT: NEUTEBOON, Saskia  
; APPLICANT: SHEPARD, Michael  
; TITLE OF INVENTION: VIRAL ENZYME ACTIVATED PROTOXOPHORES

; TITLE OF INVENTION: AND USE OF SAME TO TREAT VIRAL INFECTIONS  
; FILE REFERENCE: NB 2021.00  
; CURRENT APPLICATION NUMBER: US/10/133,133A  
; CURRENT FILING DATE: 2002-04-26  
; PRIOR APPLICATION NUMBER: 60/286,983  
; PRIOR FILING DATE: 2001-04-27  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 612  
; TYPE: DNA  
; ORGANISM: Hepatitis C Virus  
US-10-133-133A-6

Alignment Scores:  
Pred. No.: 1 13e-95 Length: 612  
Score: 904.50 Matches: 178  
Percent Similarity: 92.82% Conservative: 3  
Best Local Similarity: 91.28% Mismatches: 11  
Query Match: 87.65% Indels: 3  
DB: 14 Gaps: 1

US-09-965-594-16 (1-197) x US-10-133-133A-6 (1-612)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 19 GGTAGTGTGTCATTGGGTAGGATCATTTTCCGGTAGTGTAGTATCATCGCGGTAC 78  
Qy 22 AlaGlnGlnThrArgGlyGluGlyGlyGlnGluThrSerGlnThrGlyArgAspLys 41  
Db 79 GCCACGACAGACAGGGGCTCTTAGGGTGCATATACACAGGCTAACTGGCGGACAAA 138  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 139 AACCAAGTCAGGGTGAGGTCTCAGATGTGTCAACTGCTGCCCAAAACCTTCCTGGCAACG 198  
Qy 62 CysIleAsnGlyValCysTrpThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 199 TGCATCAATGGGGTGTCTGGAGTGTCTACCGGGGCGGACAGGACCATCGCGTCA 258  
Qy 82 ProLysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTrpGln 101  
Db 259 CCCAAGGTCCTGTATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCC 318  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuThrLeu 121  
Db 319 GCTTCGCAAGGTACCGGCTCATTSACACCTGCACTTGGGCTCTCGGACCTTTACCTG 378  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 379 GTCACGAGGACCGCGATGTATCCGTGCGCGGGGGGTGATAGAGGGGACCGCTG 438  
Qy 142 LeuSerProArgProIleSerThrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 439 CTGTGCGCGCGGGCCCAATTCCTACTTGAAGGCTCTCGGGGGGTCCGCTGTGTGGCCC 498  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAla 181  
Db 499 GCGGGGACCGCGTGGGCATATTAGGGCCCGGTGTGCACCCGTGGAGTGGCTAAGGCG 558  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
Db 559 GTGGACTTTATCCCTGTGGAGACCTTAGACACAACCATGAGGTCC 603

RESULT 10

US-09-742-659-3  
; Sequence 3, Application US/09742659  
; Patent No. US20010034019A1  
; GENERAL INFORMATION:  
; APPLICANT: Hong, Zhi  
; APPLICANT: Butkiewicz, Nancy J.  
; APPLICANT: Zhong, Weidong  
; APPLICANT: Ingravallo, Paul

; APPLICANT: Wright-Minoque, Jacquelyn  
; APPLICANT: Lau, Johnson Y.  
; APPLICANT: Lemon, Stanley M.  
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses  
; FILE REFERENCE: ID01116  
; CURRENT APPLICATION NUMBER: US/09/742,659  
; CURRENT FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: US 60/171,469  
; PRIOR FILING DATE: 1999-12-22  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 9646  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-742-659-3  
Alignment Scores:  
Pred. No.: 2 78e-92 Length: 9646  
Score: 888.50 Matches: 172  
Percent Similarity: 89.22% Conservative: 10  
Best Local Similarity: 84.31% Mismatches: 13  
Query Match: 86.09% Indels: 9  
DB: 1 Gaps: 1

US-09-965-594-16 (1-197) x US-09-742-659-3 (1-9646)

Qy 3 LysLysGlySerValIleValGlyArgIleAsn----- 14  
Db 3354 CGTAGGGCCAGGATGATCTGTGGACACCGCGGAATGGTCTCCAAAGGGTGGAGG 3413  
Qy 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlu 33  
Db 3414 TTGCTGGCGCCATCAGCGGTACGCGGACGACGAGAGAGGCTCTCTAGGTGCTAATC 3473  
Qy 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr 53  
Db 3474 ACCAGTCTGACTGGCCGGGACAAACCAAGTGGAGGTGAGTCCGATCGTGTCTACT 3533  
Qy 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73  
Db 3534 GCTACCCAAACCTTCCTGGCAACGTGCATCAATGGGTATGCTGTGACTGTCTACCACGG 3593  
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrThrAsnVal 93  
Db 3594 GCGGAAACGAGGACCATCGCATCACCAAGGGTCTCTATCCAGATGTATACCAATGTG 3653  
Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113  
Db 3654 GACCAAGACCTTGTGGGTGGCGGCTCCTCAAGGTTCGCGCTCATGACACCTGCACC 3713  
Qy 114 CysGlySerSerAspLeuThrLeuValThrArgHisAlaAspValIleProValArgArg 133  
Db 3714 TGCGGCTCCTCGGACCTTTTACCTGGTTACGAGGACGCGGACGCTATTCGCGCGCGG 3773  
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153  
Db 3774 CGAGGTGATAGAGGGTAGCTGCTTTCGCCCGCGGCCCATTTCTACCTAAAGGCTCC 3833  
Qy 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
Db 3834 TCGGGGGTCCGCTGTTGTGCGCGCGGACGACGCGGCTATTACAGGGCGCGGGTG 3893  
Qy 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
Db 3894 TGCACCGGTGGAGTGGCCAAAGGGGTGAGCTTTATCCCTGTGGAGAACCTAGACACAAC 3953  
Qy 194 MetArgSerPro 197  
Db 3954 ATGAGATCCCCG 3965  
RESULT 11  
US-09-238-076-1

```

; Sequence 1, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-238-076-1

Alignment Scores:
Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 10 Gaps: 1

US-09-965-594-16 (1-197) x US-09-238-076-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn-----14
Db 3354 CGTAGGGCCAGGAGATACCTGGCCAGCCGACGGAATGGTCTCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaThrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu 33
Db 3414 TTGCTGGCCGCCATCAGCGCGTACGCCAGCAGAGAGAGGCGCTCTAGGGGTGTAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCAGCGCTGACTGGCGCGGACAAAACCAAGTGGAGGTGAGTCCAGATCGTGCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
Db 3534 GCTACCAACACCTTCTGGCAACGTCATCAATGGGTATGCTGGACTGCTTACCACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTrpThrAsnVal 93
Db 3594 GCGGGAACGAGGNGCCATCCGATCCACCAAGGGTCTGTGCATCCAGATGTATACCAATGTG 3653

```

```

QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGGGTGGCCCTCTCAAGGTTCCCGCTCATGACACCTGACCC 3713
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGCGCTCTCGGACCTTACCTGTGTACGAGGACGCGCATGTCTATTCCTCCGTCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTrpLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTTCGCCCCGCCCATTTCTACTTGAAGGGTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGTCCGCTGTTGTGCCCGCGGACACGCGTGGGCTATTACAGGCGCGGTG 3893
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3894 TGCACCCGTTGGAGTGGCTAAGCGGTGGACTTTATCCCTGTGGAGACCTAGACACACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCG 3965

RESULT 12
US-09-995-937-1
; Sequence 1, Application US/0995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-May-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-995-937-1
Alignment Scores:

```

```
Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-995-937-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATCTGCTTGGCCACGCGCAAGTGTCTCCAAAGGGGTGGAGG 3413
QY 15 --LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlu 33
DB 3414 TTGCTGGCCCATCAGCGGTGAGCCAGCAGAGAGGCTCTAGGGTGTATAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlnValGlnIleValSerThr 53
DB 3474 ACCAGCCTGACTGGCCGGGACAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
DB 3534 GCTACCCAAACCTTCCTGGCAACGTGATCAATGGGTATGCTGGACTGTCTACCAAGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
DB 3594 GCGGACGAGGACCATCGCATCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCCGCTCCCAAGTTCCCGCTATGTACACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGGCTCTCGGACCTTACCTGGTCACGAGCAGCGCATCTCCCTGCGCCCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
DB 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGGCGCATTTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
DB 3834 TCGGGGGTCCGCTGTGTGGCCCGGGACACCGCTGGGCTATTCAGGGCCGGGTG 3893
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
DB 3894 TGCACCCGTGGAGTGGCTAAGGGCGGTGGACTTTATCCCTGTGGAGAACCCTAGAGACAAC 3953
QY 194 MetArgSerPro 197
DB 3954 ATGAGATCCCGC 3965

RESULT 13
US-09-917-563-1
; Sequence 1, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
;
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/917,563
FILING DATE: 27-Jul-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 09/238,076
FILING DATE: 26-JAN-1999
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9646 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-917-563-1

Alignment Scores:
Pred. No.: 2,78e-92 Length: 9646
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-917-563-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATCTGCTTGGCCACGCGCAAGTGTCTCCAAAGGGGTGGAGG 3413
QY 15 --LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlu 33
DB 3414 TTGCTGGCCCATCAGCGGTGAGCCAGCAGAGAGGCTCTAGGGTGTATAATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlnValGlnIleValSerThr 53
DB 3474 ACCAGCCTGACTGGCCGGGACAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValTyrHisGly 73
DB 3534 GCTACCCAAACCTTCCTGGCAACGTGATCAATGGGTATGCTGGACTGTCTACCAAGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
DB 3594 GCGGACGAGGACCATCGCATCCCAAGGGTCTGTCTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCCGCTCCCAAGTTCCCGCTATGTACACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGGCTCTCGGACCTTACCTGGTCACGAGCAGCGCATCTCCCTGCGCCCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
DB 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGGCGCATTTCTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
DB 3834 TCGGGGGTCCGCTGTGTGGCCCGGGACACCGCTGGGCTATTCAGGGCCGGGTG 3893
```



QY 174 CysThrArgGlyValAlaValAspPheIleProValGluSerLeuGluThr 193  
|||||  
Db 3894 TGCACCGTGGAGTGGCTAAGCGGTGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3953  
|||||  
QY 194 MetArgSerPro 197  
|||||  
Db 3954 ATGAGATCCCG 3965  
|||||  
RESULT 14  
US-09-238-076-5  
; Sequence 5, Application US/09238076  
; Patent No. US20020102540A1  
; GENERAL INFORMATION:  
; APPLICANT: RICE, CHARLES et al.  
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HOWELL & HAERKAMP, L.C.  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MO  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/238.076  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/034,756  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 6029-4831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 314-727-5188  
; TELEFAX: 314-727-6092  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12980 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-09-238-076-5  
Alignment Scores:  
Pred. No.: 4,06e-92 Length: 12980  
Score: 888.50 Matches: 172  
Percent Similarity: 89.22% Conservative: 10  
Best Local Similarity: 84.31% Mismatches: 13  
Query Match: 86.09% Indels: 9  
DB: 10 Gaps: 1  
US-09-965-594-16 (1-197) x US-09-238-076-5 (1-12980)  
QY 3 LysLysGlySerValIleValGlyArgIleAsn----- 14  
Db 3354 CGTAGGGCCAGAGATAGTGTGGCCAGCGGAGTGTCTCAAGGGGTGGAGG 3413  
|||||  
QY 15 ---LeuSerGlyAspThrAlaThrArgGlnGlnThrArgGlyGluGluCysGlnGlu 33  
|||||  
Db 3414 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGCAGAGAGCGCTCTAGGTGTATATC 3473  
|||||  
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53  
|||||

Db 3474 ACAGCGTCACTGGCGGGGACAAACCAAGTGGAGGTGAGGTCCAGATCGTGTCAACT 3533  
|||||  
QY 54 AlaThrGlnIlePheLeuAlaThrCysIleAsnGlyValCysIlePheValThrHisGly 73  
|||||  
Db 3534 GCTAGCCAAACCTTCTGTGGCAACGTGCATCAATGGGTATGCTGACATGCTACCAACGG 3593  
|||||  
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrAsnVal 93  
|||||  
Db 3594 GCGGGAACGAGGACCATCGATCAACCCAGGGTCTGTCTATCCAGATGATACCAATGTG 3653  
|||||  
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113  
|||||  
Db 3654 GACCAAGACCTTGTGGGTGGCGCGCTCTCAAGTTCCTCGCTCATTCACACCTGCACACC 3713  
|||||  
QY 114 CysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArgArg 133  
|||||  
Db 3714 TGGCGTCTCTCGGACCTTTACCTGTGTACGAGGACGCGGATGTCATTCCTGCGCGCGG 3773  
|||||  
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153  
|||||  
Db 3774 CGAGGTGATAGCAGGGTAGCTGCTTTCGCCCGCGCCCATTTCTTCTACTTGAAGGCTCC 3833  
|||||  
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173  
|||||  
Db 3834 TCGGGGGTCCGCTGTTGTGCCCGCGGACACGCGGTGGGCTATTTCAGGCGCGCGGTG 3893  
|||||  
QY 174 CysThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThr 193  
|||||  
Db 3894 TGCACCGTGGAGTGGCTAAGCGGTGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3953  
|||||  
QY 194 MetArgSerPro 197  
|||||  
Db 3954 ATGAGATCCCG 3965  
|||||  
RESULT 15  
US-09-995-937-5  
; Sequence 5, Application US/09995937  
; Publication No. US20030028010A1  
; GENERAL INFORMATION:  
; APPLICANT: RICE, CHARLES et al.  
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HOWELL & HAERKAMP, L.C.  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MO  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/995,937  
; FILING DATE: 28-May-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/034,756  
; FILING DATE: 04-May-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 6029-4831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 314-727-5188  
; TELEFAX: 314-727-6092  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12980 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-937-5

Alignment Scores:
Pred. No.: 4.06e-92 Length: 12980
Score: 888.50 Matches: 172
Percent Similarity: 89.22% Conservative: 10
Best Local Similarity: 84.31% Mismatches: 13
Query Match: 86.09% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-16 (1-197) x US-09-995-937-5 (1-12980)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
DB 3354 CGTAGGGCCAGGAGATACTGCTGGCCGACGCGGAATGGTCTCCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaIleValGlnGlnThrArgGlyGluGluCysGlnGlu 33
DB 3414 TTGCTGGGCCCATCATCGGGGTACGCCACGACGACGAGAGGCTCTTAGGGTGTATATC 3473
QY 34 ThrSerGlnThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
DB 3474 ACCAGCTGACTGGCCGGGACAAACCAAGTGGAGGGTGAGGTCCAGATCGTGTCACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThrValThrHisGly 73
DB 3534 GCTACCCAAACCTTCCTGGCAACGTGCATCAATGGGGTATGCTGGACTGCTACCAACGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetThrThrAsnVal 93
DB 3594 GCGGGAACGAGGACCATCGCATCACCAGGGTCTGTATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
DB 3654 GACCAAGACCTTGTGGCTGGCCCGCTCTCTCAAGGTTCGGCTCATTTGACACCTGCACC 3713
QY 114 CysGlySerSerAspLeuThrLeuValThrArgHisAlaAspValIleProValArgArg 133
DB 3714 TGGGGCTCTCGGACCTTTACTCTGGTCACGAGGACGCGCATGCTATCCCGTGGCGCG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerThrLeuLysGlySer 153
DB 3774 CGAGTGATAGCAGGGGTAGCTGTCTTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaIleVal 173
DB 3834 TCGGGGGTCCGTGTGTGTCGCCCGGGACACGCGGTGGCGCTATTTCAGGGCGCGGTG 3893
QY 174 CysThrArgGlyValAlaValAlaValAspPheIleProValGluSerLeuThrThr 193
DB 3894 TGCACCCGTGGAGTGGCTAAGCGGGTGGACTTTATCCCTCTGGAGAACCTTAGACAACC 3953
QY 194 MetArgSerPro 197
DB 3954 ATGAGATCCCCG 3965
```

Search completed: August 31, 2003, 04:54:21  
Job time : 190.482 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-16  
Perfect score: 1032  
Sequence: 1 MKKKGSVVIGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-O=/cpn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151919\_28322/app\_query.fasta\_1.2872  
-DB=EST -QVMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LORPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGCN\_1.1\_12630\_@runat\_29082003\_151919\_28322 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estab:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	101.5	9.8	1403	13	BQ926101	BQ926101 AGENCOURT
c 2	99.5	9.6	779	10	BF631437	BF631437 HVSMEB001
c 3	98.5	9.5	1199	13	BQ892487	BQ892487 AGENCOURT
c 4	98	9.5	846	10	BF182274	BF182274 601804028
c 5	98	9.5	984	10	BF304699	BF304699 601888252
c 6	97.5	9.4	905	13	BQ542842	BQ542842 AGENCOURT
c 7	97	9.4	1141	11	AK080545	AK080545 Mus muscu
c 8	96	9.3	615	12	BQ001625	BQ001625 BJ001625
c 9	96	9.3	643	12	BQ024121	BQ024121 BJ024121
c 10	96	9.3	754	12	BJ016176	BJ016176 BJ016176
c 11	96	9.3	1031	14	CB950999	CB950999 AGENCOURT
c 12	95	9.2	580	14	CA728398	CA728398 wdlc.pk0
c 13	95	9.2	961	10	BF203316	BF203316 601865914
c 14	95	9.2	1146	12	BM915803	BM915803 AGENCOURT
c 15	94	9.1	701	10	BF863244	BF863244 963042C02
c 16	94	9.1	1100	10	BG420390	BG420390 602452419
c 17	94	9.1	1101	29	BZ567280	BZ567280 pacas2-164
c 18	93.5	9.1	898	14	CA787713	CA787713 AGENCOURT
c 19	93.5	9.1	1505	10	BF183416	BF183416 601809557
c 20	93	9.0	644	29	BX238988	BX238988 Danio rer
c 21	93	9.0	1411	11	BC020343	BC020343 Homo sapi
c 22	92.5	9.0	488	10	BF776637	BF776637 287489 NA
c 23	92.5	9.0	993	9	AL555424	AL555424 AL555424
c 24	92.5	9.0	1329	13	BQ960995	BQ960995 AGENCOURT
c 25	92	8.9	649	10	BE289911	BE289911 601089126
c 26	92	8.9	757	12	BI258851	BI258851 602969666
c 27	91.5	8.9	502	9	AA036834	AA036834 zk29405.r
c 28	91.5	8.9	539	10	BE757615	BE757615 212104 MA
c 29	91.5	8.9	844	11	CNS0904S	BN053096 Single re
c 30	91.5	8.9	844	12	BI198486	BI198486 602760491
c 31	91.5	8.9	938	13	BQ894657	BQ894657 AGENCOURT
c 32	91.5	8.9	1000	13	BQ735135	BQ735135 AGENCOURT
c 33	91	8.8	470	13	BQ758584	BQ758584 EBma07_SO
c 34	91	8.8	471	13	BU978992	BU978992 HA14N15r
c 35	91	8.8	515	14	CA023748	CA023748 HZ47E17r
c 36	91	8.8	583	12	BM374064	BM374064 HVSMEC001
c 37	91	8.8	938	10	BG309750	BG309750 HVSMEC001
c 38	91	8.8	1283	13	BQ709745	BQ709745 AGENCOURT
c 39	90.5	8.8	528	28	AQ620249	AQ620249 HS_5182_B
c 40	90.5	8.8	577	29	BZ904625	BZ904625 CH240_261
c 41	90.5	8.8	2904	11	AK046602	AK046602 Mus muscu
c 42	90	8.7	460	14	CB883286	CB883286 HQ01M02w
c 43	90	8.7	500	12	BM708007	BM708007 UI-E-C11-
c 44	90	8.7	553	9	AL856435	AL856435 AL856435
c 45	90	8.7	569	12	BM825317	BM825317 K-EST0097

ALIGNMENTS

RESULT 1  
BO926101/c  
LOCUS BO926101 1403 bp mRNA linear EST 20-AUG-2002  
DEFINITION AGENCOURT\_8752655 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6335718  
5', mRNA sequence.  
ACCESSION BO926101  
VERSION BO926101.1 GI:22341132  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1403)



ИЗДАНИЕ СООБЩАЕТ ЗАКАЗЧИК

BF182274 846 bp mRNA linear EST 31-OCT-2000  
601804028F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035102 5',  
mRNA sequence.

```

VERSION      BF182274.1  GI:11060416
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM
REFERENCE    1 (bases 1 to 846)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM9308 row: g column: 07
            High quality sequence stop: 696.
            Location/Qualifiers
                1..846
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="IMAGE:4035102"
                /tissue_type="tumor, gross tissue"
                /dev_stage="7 months"
                /lab_host="DH10B"
                /clone_lib="NCI-CGAP_Mam5"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
                Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Robin Humphreys,
                NIH"
BASE COUNT   176 a 218 c 241 g 210 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:    17.6      Length:    846
Score:        98.00     Matches:   46
Percent Similarity: 48.00%  Conservative: 14
Best Local Similarity: 36.80%  Mismatches: 35
Query Match:  9.50%    Indels:   31
DB:           10       Gaps:     9
US-09-965-594-16 (1-197) x BF182274 (1-846)
QY 73 GlyAlaGlyThrArg-ThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAs 92
DB 757 GGTCTGCTACCAACAGCGCTGGATGACAGAAAGGACCA-----CATCTTC 710
QY 92 nValAspLysAspLeuValGlyTrpGln-----AlaProG1 104
DB 709 GGTCTTCAGTCCCAAGTGGCTGGAGAAAGTGAACACAGAGAGAGGACGTCCTCA 650
QY 104 nGlySerArg-----SerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVa 122
DB 649 GTCCCCAGCGTTCAGTGTAGTACAAAGTGTCTGCTGGA-----610
QY 122 lThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeuLeu 142
DB 609 -ACTAGACACACCT--GTAATCCAGGAGGAACGCGTGGAGAACACAGAGGACCTCC--CT 556
QY 142 UserProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu---LeuCysPr 161
DB 555 CACCCCACTCC---TCCGCTCAGCGGACACCTCTCTCTGGCCCACTCCCTCTGTC 499
QY 161 oAlaGlyHis---AlaValGlyIlePheArg-----AlaAlaValCysThrAr 176
DB 498 TAGTGGGACCTCTCTCCCAAGCCACACGACTGTTACTCCCTTTGGCCCTCTGCACTCT 439

```

---

```

QY 176 gGlyValAlaLys 180
DB 438 TGGGATGACTGAG 426
RESULT 5
BF304699/c
LOCUS      601888252f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:412276 5',
DEFINITION mRNA sequence.
ACCESSION  BF304699
VERSION    BF304699.1  GI:11251586
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 984)
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS   Unpublished
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L1CM1005 row: g column: 13
            High quality sequence stop: 646.
            Location/Qualifiers
                1..984
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:412276"
                /tissue_type="rhabdomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_17"
                /note="Organ: muscle; Vector: pOTB7; Site.1: EcoRI;
                Site.2: XhoI; cDNA made by oligo-dt priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCAGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
```

---

```

BASE COUNT   133 a 329 c 351 g 171 t
ORIGIN
Alignment Scores:
Pred. No.:    21.4      Length:    984
Score:        98.00     Matches:   32
Percent Similarity: 44.05%  Conservative: 5
Best Local Similarity: 38.10%  Mismatches: 25
Query Match:  9.50%    Indels:   22
DB:           10       Gaps:     5
US-09-965-594-16 (1-197) x BF304699 (1-984)
QY 100 TrpGlnAlaProGlnGlySerArgSerLeuThr---ProCysThrCysGlySerSerAsp 118
DB 646 TGGCCCACTCCAGCGCATTCCTCGGAGAGAGGACCGGTGACCTGC-----599
QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
DB 598 -----ACCAGGACGCGGACACATACATCAAGGAGAGCGTGT---TCCCGC 554
QY 139 GlySerLeuSerProArgPro-----IleSerTyrLeuLysGlySer 153
DB 553 GGGCGCCTCTTGTGGGAGAGACCTCGATGTGTGTCCAAAGTCGCGTGTACTGGAAGT 494

```

```

QY 154 SerGlyCysProLeuLeuCysProAlaGlyHisAlaValGlyLePheArgAlaVal 173
Db 493 CGCAGCGCTCCGTCAGTGCAGC-----TTCCAGCGCCCGGG 455
QY 174 CysThrArgGly 177
Db 454 TGCGCCGAGGA 443

RESULT 6
BU542842/c
LOCUS BU542842 905 bp mRNA linear EST 13-SEP-2002
DEFINITION AGENCOURT_10334715 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574789
S. mRNA sequence.
ACCESSION BU542842
VERSION BU542842.1 GI:22853325
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 905)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2770 row: k column: 13
High quality sequence stop: 633.
Location/Qualifiers
1..905
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6574789"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 201 a 260 c 273 g 171 t
ORIGIN
Alignment Scores:
Pred. No.: 21.5 Length: 905
Score: 97.50 Matches: 51
Percent Similarity: 31.40% Conservative: 14
Best Local Similarity: 24.64% Mismatches: 75
Query Match: 9.45% Indels: 67
DB: 13 Gaps: 9

US-09-965-594-16 (1-197) x BU542842 (1-905)

QY 28 GluGluGlyCysGlnGluThrSerGln---ThrGlyArgAspLysAsnGlnValGluGly 46
Db 884 AAGAGAGGCCCGCCAGTCTGTTCCCGAGGAAGGGGACCCGAGACCAAGAGGAGGAGGC 825
QY 47 GluValGlnIleValSerThrAlaThrGlnThrPheLeuAla----- 60
Db 824 GGGGCCCTTCTCCAGGCCCTGTCACAAAGTGTCCTTGGGTGGCCCGCCATGGTCCCA 765

```

```

QY 61 -----ThrCysIleAsnGlyValCys----- 67
Db 764 CATTTCTGCAGCATCCGGCAGAACATGTGTGGGTCTTGCCTCCAGCAGCAGGACAGCC 705
QY 68 ---TrpThrValTyrHisGlyAla----- 74
Db 704 AAGTGGGAGGAGGAGGATGTGTGCACACAGCTGGGAGGAGGCCCTGGTGAGAAGCAGCCCA 645
QY 75 -----GlyThrArgThrIle 79
Db 644 CAGTAGACGCCCATCCAGAGGAGAACCATCTCCGAGGGCCACAGGCCCTCTGCAGCCCTG 585
QY 80 AlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGly 99
Db 584 GCACCTGCCGCCAGCCCTCCATCTCAGCGGATGTCAGAGGTGAGACAGCAATCCAGGA 525
QY 100 TrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu 119
Db 524 GTTCTGCCCTAGGTGAGCCTCTTCATCCGCCCTGTTGCTGCTGATGCTCAAGGTG 465
QY 120 -----TyrLeuValThrArgHisAlaAsp-----ValIleProValArg 133
Db 464 CCCGTCTCCACAGCTGTCGCAACGCCATCCAGGGCTTCGTCTTGTCTCTCCAGCTCACT 405
QY 134 Arg-----GlyAspSerArgGlySerLeuLeuSerPro-----Arg 145
Db 404 CGGCTCCAGGCCAGGCCCTTCATCTCTCAGGATCTGGGTTAGTTCTCTGGGTATCTG 345
QY 146 ProIleSerTyrLeuLysGlySerSerGlyClyProLeuLeuCysProAlaGly----- 163
Db 344 CCTCAGAAAGGGCTGGCAGGCTTGTCTGACAGGTGACGTGCTGCTGCTGCTGCTGCT 285
QY 164 -----HisAlaValGly 167
Db 284 CGGTGGCTCAGCGGTGCAGG 264

RESULT 7
AK080545 1141 bp mRNA linear HTC 05-DEC-2002
LOCUS AK080545
DEFINITION Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:A730082L10 product:weakly similar to zinc
finger protein (fragment) [Mus musculus], full insert sequence.
ACCESSION AK080545
VERSION AK080545.1 GI:26348600
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Alizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

```

```

JOURNAL      Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE      20530913
PUBMED      11076861
REFERENCE    4
AUTHORS      Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
              Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
              Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
              Saito,T., Okazaki,Y., Goshori,T., Bono,H., Kasukawa,T., Saito,R.,
              Kadota,K., Matsuda,H., Ashtburner,M., Batalov,S., Casavant,T.,
              Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
              Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
              Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M.,
              Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
              Balarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
              Carninci,P., de Bernaldo,M.F., Brownstein,M.J., Bult,C., Hill,D.,
              Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
              Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
              Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
              Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
              Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H.,
              Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
              Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
              and Hayashizaki,Y.
TITLE        Functional annotation of a full-length mouse cDNA collection
JOURNAL      Nature 409 (6821), 685-690 (2001)
MEDLINE      21085660
PUBMED      11217851
REFERENCE    5
AUTHORS      The FANTOM Consortium and the RIKEN Genome Exploration Research
              Group Phase I & II Team.
TITLE        Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
JOURNAL      Nature 420, 563-573 (2002)
MEDLINE      6 (bases 1 to 1141)
PUBMED      12127851
REFERENCE    5
AUTHORS      Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
              Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
              Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
              Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
              Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
              Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
              Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
              Okazaki,Y., Saito,R., Shibata,K., Sakai,C., Sakai,K., Sakazume,M.,
              Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
              Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S.,
              Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
              Muramatsu,M. and Hayashizaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
              Physical and Chemical Research (RIKEN), Laboratory for Genome
              Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
              RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
              Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,
              URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
              Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
              Encyclopedia Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in RIKEN.
              Division of Experimental Animal Research in Riken contributed to
              prepare mouse tissues.
              Please visit our web site for further details.
              URL:http://genome.gsc.riken.go.jp/
              URL:http://fantom.gsc.riken.go.jp/.
FEATURES     source
              Location/Qualifiers
                1..1141
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM DB:A730082L10"
                /db_xref="taxon:10090"
                /clone="A730082L10"
                /tissue_type="cerebellum"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="7 days neonate"

```

---

```

<1..587
/note="unnamed protein product; putative
weakly similar to zinc finger protein (fragment) [Mus
musculus] (PIR|I48722, evidence: FASTY, 50.7%ID,
57.6%length, match=601)"
/codon_start=3
/db_xref="GI:26348601"
/translation="DLSLPASPCRSLLTPRGDGFLEKLSARAVGPGSPVAFVS
TVRGAQAGCGRVRGRSEGLSKRPFRHVPVPGVHTGLSGRRIPPPAGE
AAAGRAQQVPHPPGPHGTVPVPOGAAGLLPALAAQVPGVGRGREGPRAPHS
PKPVPTALGFSFGGGGAPPILLAPANGRSVGLAL"
polyA_signal
1118..1123
/note="putative"
1141
/note="putative"
244 a 316 c 353 g 228 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 32.5 Length: 1141
Score: 97.00 Matches: 46
Percent Similarity: 39.31% Conservative: 11
Best Local Similarity: 31.72% Mismatches: 53
Query Match: 9.40% Indels: 35
DB: 11 Gaps: 8
US-09-965-594-16 (1-197) x AK080545 (1-1141)
QY 68 TTPThrValTyrHisGlyAlaGly-----ThrArgThrIleAla 80
Db 278 TGGCAGACACATCCCGCGCGCGGAGAGCGGAGCGAGCGCGCGCGCAGCA 337
QY 81 Ser-ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTr 100
Db 338 AGTGGCGCATCCCGCTGGCGGCCACATGGAACAGTGTG-----377
QY 100 pGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTy 120
Db 378 ---CCTCTCAAGGAGCGGCTGGGCTCTCTCCCTCA-----410
QY 120 rLeuValThrArgHisAlaAspValIleProValArg---ArgArgGlyAspSerArgG1 139
Db 411 -CTCGCAGCTCGCCCAAGTCTCTGTGGCGGTTAGGCGCGGAGGACCAAGAGGCGC 469
QY 139 ySerLeuLeuSerProArgProile-----SerTyrLeuLysGlyse 153
Db 470 ACCGAGACACACGCCCAAGCGGTCCTACAGCCTTGGGTTTCTGTTGGCAGGTTGG 529
QY 153 rSerGlyGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 172
Db 530 GCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 586
QY 172 aValCysThrArgGlyValAlaAlaLysAlaVal-----AspPheIleProValGluSerLe 190
Db 587 AGTGTGTGAGTGGGAGACTTGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 646
QY 190 uGluThrThrMet 194
Db 647 GCGAGGTACACTT 659
RESULT 8
BJ001625/c
LOCUS BJ001625 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 5',
DEFINITION mRNA sequence.
ACCESSION BJ001625
VERSION BJ001625.1 GI:17364516
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```





QY 162 AlaGlyHisala 165  
 DB 611 GCCTCTGCTGCA 622

RESULT 10  
 BJO16176  
 LOCUS BJO16176 754 bp mRNA linear EST 05-DEC-2001  
 DEFINITION BJO16176 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA025C02 3',  
 mRNA sequence.  
 ACCESSION BJO16176  
 VERSION BJO16176.1 GI:17376695  
 KEYWORDS EST.  
 SOURCE Oryzias latipes (Japanese medaka)  
 ORGANISM Oryzias latipes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;  
 Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.  
 REFERENCE 1 (bases 1 to 754)  
 AUTHORS Kohara, Y., Shin, I., T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.  
 TITLE Medaka EST Project in Takada's lab  
 JOURNAL Unpublished  
 COMMENT Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..754  
 /organism="Oryzias latipes"  
 /mol\_type="mRNA"  
 /strain="Hd-rk"  
 /db\_xref="taxon:8090"  
 /clone="MF01SSA025C02"  
 /sex="mixture of female and male"  
 /tissue\_type="whole embryo"  
 /dev\_stage="segmentation stage 20 - 25"  
 /clone\_lib="MF01SSA cDNA"  
 BASE COUNT 194 a 181 c 181 g 198 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 23.8 Length: 754  
 Score: 96.00 Matches: 43  
 Percent Similarity: 34.72% Conservative: 7  
 Best Local Similarity: 29.86% Mismatches: 58  
 Query Match: 9.30% Indels: 36  
 DB: 12 Gaps: 6

US-09-965-594-16 (1-197) x BJO16176 (1-754)

QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 242 AAAATGACGTAGAACCAAAAGACACACAGATCCACACACATGCTCTGTTCTACGGGCT 301  
 QY 61 ThrCysIleAsnGlyValCysTriThrValThrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 302 -----TGTTGGAGAACCTATCACAGTTCCTGCTTTAGACGACGGCA 343  
 QY 81 SerProLys-----GlyProValThrGlnMetTyThrAsnValAspLys 95  
 DB 344 GCTCTGGCGGCGGAGGAGCTCTTGGGCGCAGTGTGACTCGT----- 385  
 QY 96 AspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGly 115  
 DB 386 -----GGAGGACGAAGAGCGCTCCACCCGGAGCTGTAGCTGCAGGAGTGGGTGGC 439  
 QY 116 SerSerAspLeuTyThrValThrArg----- 124  
 DB 440 TCTGCT-----TTGGTTCTCTGCTCTCTCTGATCATCTTCTCACCTGACCTTCCA 490

QY 125 HisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerPro 144  
 DB 491 CATCCAGGTGTCCGACGCGCTGTCTGACGGGTGATGGGAGAGCGCGACAGCAGCACT 550  
 QY 145 Arg-----ProIleSerTyLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 551 CGGGGGTGAATCTCTGTCAGGACGCTCTACGGCGGATCAGGAGGACCGACTCGCTGCAGA 610  
 QY 162 AlaGlyHisala 165  
 DB 611 GCCTCTGCTGCA 622

RESULT 11  
 CB950999  
 LOCUS CB950999 1031 bp mRNA linear EST 29-APR-2003  
 DEFINITION AGENCOURT\_13445496 NIH\_MGC\_177 Mus musculus cDNA clone  
 IMAGE:30316162 5', mRNA sequence.  
 ACCESSION CB950999  
 VERSION CB950999.1 GI:30205777  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1031)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: NDCM107 row: b column: 11  
 High quality sequence stop: 333.  
 Location/Qualifiers  
 1..1031  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30316162"  
 /lab\_host="DH10B (TI-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_177"  
 /note="Organ: liver; Vector: pDNR-LIB; Site\_1: Sfil  
 (ggcattatggcc); Site\_2: Sfil (ggccgctcgcc); cDNA made  
 by oligo-dT priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGTATCAGCAGGAGTGGCATTACGGCGG-3' and  
 5'-ATTCTAGAGCGGAGCGGCGGACATG-dt(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.  
 Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."

BASE COUNT 235 a 309 c 211 g 275 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 35.7 Length: 1031  
 Score: 96.00 Matches: 42  
 Percent Similarity: 42.11% Conservative: 14  
 Best Local Similarity: 31.58% Mismatches: 54  
 Query Match: 9.30% Indels: 24  
 DB: 14 Gaps: 6

US-09-965-594-16 (1-197) x CB950999 (1-1031)

QY 44 ValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrCysIle 63  
 DB 395 ATTCAGGCTATCCCAAAAGAGAGATACATCGGAGCTTTTCTTCTCACTATTG 454

US-09-965-594-16 (1-197) x CA728398 (1-580)

QY	60	AlaThrCysIleAsnGlyValCysTrpThrVal-----TyrHisGlyAlaGlyThrArg	77
Db	40	GCATGGTGTTCTTCTGCTTCTGGTGGACCGCGGGAAGGTGGGAGCAGCAAGCGG	99
QY	78	ThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeu	97
Db	100	CGCGGGGATCTGCTGCGGTGGCGGTGGCCAGGCTGGCGGACA-----	147
QY	98	ValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCys-----Gly	115
Db	148	-----TGGAGAGCGGCACCGGCTCTGCTACTCTCAGCTGACCGCGCTACCTGGC	201
QY	116	SerSerAsp-LeuTyrLeuValThrArgHisAlaAspValIleProVal-----	131
Db	202	CGGCCATCATCTGCACCTTCTCGGGGCCATGCTCAAGTCCTACCGCCTACAGGCTCT	261
QY	132	-----ArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSer	148
Db	262	ACTAGACGGCGATTGATTCGCGCGGACGAGCGAGCTCTCCACCGCGCTCCCGTTTC	319

RESULT 13  
 BF203316/c  
 LOCUS BF203316  
 DEFINITION Homo sapiens (human)  
 ACCESSION BF203316  
 VERSION 1  
 KEYWORDS Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 961)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLCM965 row: 1 column: 03  
 High quality sequence stop: 637.

FEATURES  
 Location/Qualifiers  
 1..961  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4098578"  
 /tissue\_type="rhadomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NHLMGC\_17"  
 /note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 230 a 297 c 300 g 134 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 40.8 Length: 961  
 Score: 95.00 Matches: 28  
 Percent Similarity: 55.07% Conservative: 10



## FEATURES

## SOURCE

Location/Qualifiers  
 1. .701  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="CC-1690 wild type mt+ 2lgr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, Stress condition I,  
 normalized, Lambda Zap II"  
 /note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; This library, constructed by John Davies and Jeffrey  
 McDermott, combines cDNAs from CC-1690 cells grown to  
 mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,  
 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr  
 , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was  
 purified from each sample, pooled and cDNA synthesized.  
 The cDNA was directionally cloned into lambda zap II  
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
 pBluescript II SK- plasmids were excised from the lambda  
 Zap clones by superinfection with ExAssist (Stratagene)  
 phage. The library was normalized using method 4 described  
 in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	33.9	Length:	701
Score:	94.00	Matches:	32
Percent Similarity:	40.71%	Conservative:	14
Best Local Similarity:	28.32%	Mismatches:	45
Query Match:	9.11%	Indels:	22
DB:	10	Gaps:	4

US-09-965-594-16 (1-197) x BF863244 (1-701)

Qy	71	TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal	86
		:::	::: :::
Db	171	CACCACCATACCCTTGCTCTCAGCTGCTCACCACCAAAATTATGCCCATACGGGCCACTA	230
Qy	87	ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaProGlnGlySer	106
		:::	
Db	231	ACAAAGTTACATACACGG-----AAGGACCAGCGCGCTTGGCCACCCCTTGGAGCCG	284
Qy	107	ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla	126
Db	285	AGAAGCCCGACCGTGCTCTCTGGGTCAICCGCATGCTCAATCTCCGCTATCAG	344
Qy	127	AspValIle-----ProValArgArgGlyAspSerArg-----	138
		:::	
Db	345	GAGATCAITTTGCATGTGGCTTTAGTACCCCAAGAGAGCCTGGGAGTGGGCATTTATAA	404
Qy	139	-----GlySerLeuLeuSerProArgProIleSerTyrLeu	150
		::: :::	
Db	405	GAAGGGACGGGAATTCGGTTTGGGAAAGTACGGCGCCCAAGGCTGACCAAGTGCTA	464
Qy	151	LysGlySerSerGlyGlyProLeuLeuCysProAlaGly	163
Db	465	CTCCAGGCGACGAATGGGAGCCTTTCGCGGTGTGCGGT	503

Search completed: August 31, 2003, 04:27:34  
 Job time : 1916.31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDHPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1017	100.0	197	21	AA15223
2	1005	98.8	197	21	AA15224
3	1002	98.5	197	21	AA15222
4	985	97.8	197	21	AA15225
5	985	96.9	197	21	AA15221
6	951	93.5	195	21	AA15220
7	946	93.0	197	21	AA15222
8	912	89.7	195	21	AA15216
9	881.5	86.7	665	20	AA1524943
					HCV NS4A-NS3 compl

10	878.5	86.4	665	20	AA1524947	HCV NS4A-NS3 compl
11	877.5	86.3	665	20	AA1524941	HCV NS4A-NS3 compl
12	877.5	86.3	665	20	AA1524942	HCV NS4A-NS3 compl
13	874.5	86.0	216	20	AA1517880	HCV NS4A-NS3 compl
14	874.5	86.0	665	20	AA1524945	HCV NS4A-NS3 compl
15	874.5	86.0	665	20	AA1524946	HCV NS4A-NS3 compl
16	873.5	85.9	665	20	AA1524940	HCV NS4A-NS3 compl
17	873.5	85.9	671	20	AA1524948	HCV NS4A-NS3 compl
18	871.5	85.7	216	20	AA1517884	HCV NS4A-NS3 compl
19	870.5	85.6	216	20	AA1517879	HCV NS4A-NS3 compl
20	870.5	85.6	216	20	AA1517878	HCV NS4A-NS3 compl
21	870.5	85.6	665	20	AA1524944	HCV NS4A-NS3 compl
22	870.5	85.6	671	20	AA1524949	HCV NS4A-NS3 compl
23	870.5	85.5	215	20	AA1517890	HCV NS4A-NS3 compl
24	867.5	85.3	216	20	AA1517882	HCV NS4A-NS3 compl
25	867.5	85.3	216	20	AA1517883	HCV NS4A-NS3 compl
26	867.5	85.3	216	20	AA1517886	HCV NS4A-NS3 compl
27	866.5	85.2	216	20	AA1517877	HCV NS4A-NS3 compl
28	864	85.0	215	20	AA1517887	HCV NS4A-NS3 compl
29	863.5	84.9	216	20	AA1517881	HCV NS4A-NS3 compl
30	863.5	84.9	216	20	AA1517885	HCV NS4A-NS3 compl
31	859	84.5	213	20	AA1517888	HCV NS4A-NS3 compl
32	859	84.5	631	20	AAW93482	HCV NS3 protein.
33	858.5	84.4	191	21	AA1544728	Hepatitis C virus
34	858.5	84.4	3011	19	AAW77397	Hepatitis C virus
35	858.5	84.4	3011	24	ABP71460	Amino acid sequenc
36	858.5	84.4	3012	23	AAU99289	Hepatitis C virus
37	855.5	84.1	3011	14	AAAR40120	HCV genomic amino
38	854.5	84.0	687	16	AAAR79223	pHCV150-encoded se
39	854.5	84.0	1648	16	AAAR79221	pHCV176-encoded se
40	854.5	84.0	1766	10	AAAP92041	Sequence encoded i
41	854.5	84.0	1786	10	AAAP90158	Protein sequence o
42	854.5	84.0	2261	10	AAAP00164	Peptide encoded by
43	854.5	84.0	2301	10	AAAP92047	Sequence encoded i
44	854.5	84.0	2436	10	AAAP92050	Sequence encoded i
45	854.5	84.0	2436	10	AAAP90288	Peptide encoded by

#### ALIGNMENTS

RESULT 1  
AA15223  
ID AA15223 standard; protein; 197 AA.  
XX  
AC AA15223;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #5.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutain.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN W0200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI; 2000-465976/40.  
XX  
DR N-PSDB; AAA73332.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX  
 PS Claim 23; Fig 15; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 XX  
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1017; DB 21; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-96;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKGGVIVGRINLSGDTAYAAQOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGGVIVGRINLSGDTAYAAQOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120  
 DB 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2  
 AAB15224  
 ID AAB15224 standard; protein: 197 AA.

XX AAB15224;  
 XX  
 XX 19-DEC-2000 (first entry)  
 XX  
 XX Hepatitis C virus NS4A-NS3 fusion protease #6.  
 XX  
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX liver failure; liver cancer; mutant; mutein.  
 XX  
 XX Hepatitis C virus.  
 XX Synthetic.  
 XX  
 XX WO200040707-A1.  
 XX  
 XX 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX  
 XX 08-JAN-1999; 99US-0115271.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 XX WPI: 2000-465976/40.  
 XX  
 XX N-PSDB; AAA73333.  
 XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX  
 PS Claim 23; Fig 16; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 XX  
 SQ Sequence 197 AA;

Query Match 98.8%; Score 1005; DB 21; Length 197;  
 Best Local Similarity 98.5%; Pred. No. 2.7e-95;  
 Matches 194; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MKKGGVIVGRINLSGDTAYAAQOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGGVIVGRINLSGDTAYAAQOTRGECCQTSOTGRDKNOVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120  
 DB 61 TSINGVLWTVYHGAGTETIASPKGPVTOMTYNDKDLVGWQAPQGSRLTPTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYILKSGSGGPLLCPCAGHAYGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3  
 AAB15222  
 ID AAB15222 standard; protein: 197 AA.

XX AAB15222;  
 XX  
 XX 19-DEC-2000 (first entry)  
 XX  
 XX Hepatitis C virus NS4A-NS3 fusion protease #4.  
 XX  
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 XX liver failure; liver cancer; mutant; mutein.  
 XX  
 XX Hepatitis C virus.  
 XX Synthetic.  
 XX  
 XX WO200040707-A1.  
 XX  
 XX 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX  
 XX 08-JAN-1999; 99US-0115271.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 XX WPI: 2000-465976/40.  
 XX  
 XX N-PSDB; AAA73331.  
 XX

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 PS Claim 23; Fig 14; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 98.5%; Score 1002; DB 21; Length 197;  
 Best Local Similarity 98.5%; Pred. No. 5.6e-95;  
 Matches 194; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVWQAPQGSRLTPTCTCGSSDLY 120  
 DB 61 TCINGVCMVTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVWQAPQGSRLTPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGGPLLCFAGHAGVGFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGGPLLCFAGHAGVGFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 4  
 AAB15225  
 ID AAB15225 standard; protein; 197 AA.  
 AC AAB15225;  
 DT 19-DEC-2000 (first entry)  
 DE Hepatitis C virus NS4A-NS3 fusion protease #7.  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutain.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX WO2000040707-A1.  
 PN 13-JUL-2000.  
 PD 06-JAN-2000; 2000WO-US00345.  
 PF 08-JAN-1999; 99US-0115271.  
 PR (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR

DR N-PSDB: AAA73334.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 PS Claim 23; Fig 17; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 97.8%; Score 995; DB 21; Length 197;  
 Best Local Similarity 98.0%; Pred. No. 2.9e-94;  
 Matches 193; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVIVGRINLSGDTAYAAQOTRGECCQETSQTGRDKNQVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVWQAPQGSRLTPTCTCGSSDLY 120  
 DB 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTYTNVDKDLVGVWQAPQGSRLTPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGGPLLCFAGHAGVGFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGGPLLCFAGHAGVGFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 5  
 AAB15221  
 ID AAB15221 standard; protein; 197 AA.  
 AC AAB15221;  
 DT 19-DEC-2000 (first entry)  
 DE Hepatitis C virus NS4A-NS3 fusion protease #3.  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutain.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX WO2000040707-A1.  
 PN 13-JUL-2000.  
 PD 06-JAN-2000; 2000WO-US00345.  
 PF 08-JAN-1999; 99US-0115271.  
 PR (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR



DR WPI: 2000-465976/40.  
 DR N-PSDB; AAA73330.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX  
 XX  
 PS Claim 23; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA;  
 Query Match 96.9%; Score 985; DB 21; Length 197;  
 Best Local Similarity 97.0%; Pred. No. 3.2e-93;  
 Matches 191; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSTQGRKNQVEGEVQIVSTATOTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSTQGRKNQVEGEVQIVSTATOTFLA 60  
 QY 61 TSINGVLTVYHGAGTRTIASPKGPVTQMTNVDKDLVGWPAQGSRSLLTPTCTGSSDLY 120  
 DB 61 TCINGVCTVYHGAGTRTIASPKGPVTQMTNVDKDLVGWPAQGSRSLLTPTCTGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGGGPGLLCPAGHAVGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGGGPGLLCPAGHAVGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 6  
 AAB15220  
 ID AAB15220 standard; protein; 195 AA.  
 XX  
 AC AAB15220;  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX  
 DR WPI: 2000-465976/40.  
 DR N-PSDB; AAA73329.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX  
 XX  
 PS Claim 23; Fig 12; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 195 AA;  
 Query Match 93.5%; Score 951; DB 21; Length 195;  
 Best Local Similarity 94.9%; Pred. No. 1e-89;  
 Matches 187; Conservative 1; Mismatches 7; Indels 2; Gaps 1;  
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSTQGRKNQVEGEVQIVSTATOTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGECCQETSTQGRKNQVEGEVQIVSTATOTFLA 58  
 QY 61 TSINGVLTVYHGAGTRTIASPKGPVTQMTNVDKDLVGWPAQGSRSLLTPTCTGSSDLY 120  
 DB 59 TCINGVCTVYHGAGTRTIASPKGPVTQMTNVDKDLVGWPAQGSRSLLTPTCTGSSDLY 118  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGGGPGLLCPAGHAVGIFRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGGGPGLLCPAGHAVGIFRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195  
 RESULT 7  
 AAB15226  
 ID AAB15226 standard; protein; 197 AA.  
 XX  
 AC AAB15226;  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB; AAA73335.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX Example 5; Fig 18; 66pp; English.  
 PS  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0  
 CC wild-type sequence.  
 XX  
 SQ Sequence 197 AA;  
 Query Match 93.0%; Score 946; DB 21; Length 197;  
 Best Local Similarity 94.4%; Pred. No. 3.3e-89;  
 Matches 186; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSGTRDKNOVEGEVIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQOTRGLGCIITSLTGRDKNOVEGEVIVSTAQTFLA 60  
 QY 61 TSINGVLVTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120  
 DB 61 TCINGVCTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRRGRSGLSLSPRPISYLKSGSGGPLLCFAGHAGVIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRRGRSGLSLSPRPISYLKSGSGGPLLCFAGHAGVIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 8  
 AAB15212  
 ID AAB15212 standard; protein; 195 AA.  
 XX  
 AC AAB15212;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 PI WPI: 2000-465976/40.  
 DR N-PSDB; AAA73328.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX Example 2; Fig 10; 66pp; English.  
 PS  
 CC The present sequence is a fusion protein created using the Hepatitis C  
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
 CC essential for the replication of the virus, acting to cleave its  
 CC replicative proteins from the polyprotein produced from the HCV genome.  
 CC Inhibitors of the two proteins should be effective as antiviral  
 CC treatments of HCV infection. This is useful as HCV can lead to chronic  
 CC liver disease such as cirrhosis, liver failure and liver cancer. The  
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
 CC proteins which can be used to identify inhibitors of this type, as well  
 CC as enabling structural studies of the protease and protease-inhibitor  
 CC complexes.  
 XX  
 SQ Sequence 195 AA;  
 Query Match 89.7%; Score 912; DB 21; Length 195;  
 Best Local Similarity 92.4%; Pred. No. 1e-85;  
 Matches 182; Conservative 1; Mismatches 12; Indels 2; Gaps 1;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQOTRGEQCOETSGTRDKNOVEGEVIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRIVLNG--AYAAQOTRGLGCIITSLTGRDKNOVEGEVIVSTAQTFLA 58  
 QY 61 TSINGVLVTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 120  
 DB 59 TCINGVCTVYHAGTRTITASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLY 118  
 QY 121 LVTRHADVIPVRRRGRSGLSLSPRPISYLKSGSGGPLLCFAGHAGVIFRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRRGRSGLSLSPRPISYLKSGSGGPLLCFAGHAGVIFRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195  
 RESULT 9  
 AAY24943  
 ID AAY24943 standard; Protein; 665 AA.  
 XX  
 AC AAY24943;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex SEQ ID NO:14.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-0524528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.

XX PA (SCHE ) SCHERING CORP.  
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX PI MPI; 1999-385385/32.  
 XX DR New hepatitis C virus covalent complexes  
 XX PT Claim 6; Page 90-92; 21pp; English.  
 XX PS  
 CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.  
 XX SQ Sequence 665 AA;

Query Match 86.7%; Score 881.5; DB 20; Length 665;  
 Best Local Similarity 85.2%; Pred. No. 8e-82;  
 Matches 167; Conservative 16; Mismatches 10; Indels 3; Gaps 1;  
 QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFAT 61  
 DB 22 GSVVIVGRILILSGSGSITAYSQOTRGLLGCKKTSLTGRDKNOVEGEVQIVSTATQSFAT 81  
 QY 62 SINGVLTMTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPGQSRSLPTCTCGSSDLYL 121  
 DB 82 CVNGVCMVTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRGDSRGSLSPRPISYLGSGGGPILCPAGHAGVIFRAAVSTRGVAKA 181  
 DB 142 VTRHADVIPVRRGDSRGSLSPRPVSYLKGSGAGPILCPGSHAGVIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETTMRSP 217

RESULT 10  
 AAY24947  
 ID AAY24947 standard; Protein; 665 AA.  
 XX AC AAY24947;  
 XX DT 07-SEP-1999 (first entry)  
 XX DE HCV NS4A-NS3 complex SEQ ID NO:18.  
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 XX hydrophobic domain; covalent complex; detection; inhibitor.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN WO9928482-A2.  
 XX PD 10-JUN-1999.  
 XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.  
 XX PA (SCHE ) SCHERING CORP.

PA (SCHE ) SCHERING CORP.  
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX PI MPI; 1999-385385/32.  
 XX DR New hepatitis C virus covalent complexes  
 XX PT Claim 6; Page 100-102; 21pp; English.  
 XX PS  
 CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.  
 XX SQ Sequence 665 AA;

Query Match 86.4%; Score 878.5; DB 20; Length 665;  
 Best Local Similarity 84.7%; Pred. No. 1.6e-81;  
 Matches 166; Conservative 17; Mismatches 10; Indels 3; Gaps 1;  
 QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTRDKNOVEGEVQIVSTATQTFAT 61  
 DB 22 GSVVIVGRILILSGSGSITAYSQOTRGLLGCKKTSLTGRDKNOVEGEVQIVSTATQSFAT 81  
 QY 62 SINGVLTMTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPGQSRSLPTCTCGSSDLYL 121  
 DB 82 CVNGVCMVTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSLPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRGDSRGSLSPRPISYLGSGGGPILCPAGHAGVIFRAAVSTRGVAKA 181  
 DB 142 VTRHADVIPVRRGDSRGSLSPRPVSYLKGSGAGPILCPGSHAGVIFRAAVCTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETTMRSP 217

RESULT 11  
 AAY24941  
 ID AAY24941 standard; Protein; 665 AA.  
 XX AC AAY24941;  
 XX DT 07-SEP-1999 (first entry)  
 XX DE HCV NS4A-NS3 complex SEQ ID NO:12.  
 XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 XX hydrophobic domain; covalent complex; detection; inhibitor.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN WO9928482-A2.  
 XX PD 10-JUN-1999.  
 XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.  
 XX PA (SCHE ) SCHERING CORP.

XX  
PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
XX WPI: 1999-385385/32.

XX  
PT New hepatitis C virus covalent complexes

XX  
PS Claim 6; Page 85-87; 21pp; English.

XX  
CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity, the helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.

XX  
SQ Sequence 665 AA;

Query Match 86.3%; Score 877.5; DB 20; Length 665;  
Best Local Similarity 85.2%; Pred. No. 2.1e-81;  
Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGECCGCOETSGTGRKNOVEGEVQIVSTATOTFLAT 61  
DB 22 GSVVIVGRILSGSITAYSQOTRGLGCKITSLTGRKNOVEGEVQIVSTATQSFLAT 81  
QY 62 SINGVLMTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121  
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141  
QY 122 VTRHADYIPVRRGRDGRGSLSPRPISYLYKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKA 181  
DB 142 VTRHADYIPVRRGRDGRGSLSPRPISYLYKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKA 201  
QY 182 VDFIPVESLETTMRSP 197  
DB 202 VDFVPVESMETMRSP 217

RESULT 12

AAAY24942  
ID AAY24942 standard; Protein; 665 AA.

XX  
AC AAY24942;

XX  
DT 07-SEP-1999 (first entry)

XX  
DE HCV NS4A-NS3 complex SEQ ID NO:13.

XX  
KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX  
OS Hepatitis C virus.  
OS Synthetic.

XX  
PN W09928482-A2.

XX  
PD 10-JUN-1999.

XX  
PF 24-NOV-1998; 98WO-US24528.

XX  
PR 28-JUL-1998; 98US-0094331.

XX  
PR 28-NOV-1997; 97US-0067315.

XX  
PA (SCHE ) SCHERING CORP.

XX

PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
XX  
DR WPI: 1999-385385/32.

XX  
PT New hepatitis C virus covalent complexes

XX  
PS Claim 6; Page 88-90; 21pp; English.

XX  
CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity, the helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.

XX  
SQ Sequence 665 AA;

Query Match 86.3%; Score 877.5; DB 20; Length 665;  
Best Local Similarity 85.2%; Pred. No. 2.1e-81;  
Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSD---TAYAQOTRGECCGCOETSGTGRKNOVEGEVQIVSTATOTFLAT 61  
DB 22 GSVVIVGRILSGSITAYSQOTRGLGCKITSLTGRKNOVEGEVQIVSTATQSFLAT 81  
QY 62 SINGVLMTVYHGAGTRTITASPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121  
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141  
QY 122 VTRHADYIPVRRGRDGRGSLSPRPISYLYKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKA 181  
DB 142 VTRHADYIPVRRGRDGRGSLSPRPISYLYKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKA 201  
QY 182 VDFIPVESLETTMRSP 197  
DB 202 VDFVPVESMETMRSP 217

RESULT 13

AAAY17880  
ID AAY17880 standard; Protein; 216 AA.

XX  
AC AAY17880;

XX  
DT 07-SEP-1999 (first entry)

XX  
DE HCV NS4A-NS3 complex SEQ ID NO:4.

XX  
KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.

XX  
OS Hepatitis C virus.  
OS Synthetic.

XX  
PN W09928482-A2.

XX  
PD 10-JUN-1999.

XX  
PF 24-NOV-1998; 98WO-US24528.

XX  
PR 28-JUL-1998; 98US-0094331.

XX  
PR 28-NOV-1997; 97US-0067315.

XX  
PA (SCHE ) SCHERING CORP.

XX  
PI Malcolm BA, Taremi SS, Weber PC, Yao N;

XX DR WPI: 1999-385385/32.  
XX PT New hepatitis C virus covalent complexes  
XX PS Claim 6; Page 76-77; 21pp; English.  
XX CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity, the helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.  
XX SQ Sequence 216 AA;

Query Match 86.08; Score 874.5; DB 20; Length 216;  
Best Local Similarity 85.18; Pred. No. 8.8e-82;  
Matches 166; Conservative 16; Mismatches 10; Indels 3; Gaps 1;  
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTGRDKNQVEGEVQIVSTATQTFLAT 61  
DB 22 GSVVIVGRILLSGSGSITAYSQOTRGLGCKKTSLTGRDKNQVEGEVQIVSTATQSFAT 81  
QY 62 SINGVLWTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWQAPGSGSLTPTCTCGSSDLYL 121  
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSULTPTCTCGSSDLYL 141  
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVIFRAAVSTRGVAKA 181  
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLGKSSGGLLCPSGHAGVIFRAAVCTRGVAKA 201  
QY 182 VDFIPVESLETTMRS 196  
DB 202 VDFVPVESMETTMRSP 216

RESULT 14  
AAAY24945  
ID AAY24945 standard; Protein: 665 AA.

XX AC AAY24945;  
XX DT 07-SEP-1999 (first entry)  
XX DE HCV NS4A-NS3 complex SEQ ID NO:16.  
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.  
XX OS Hepatitis C virus.  
OS Synthetic.

XX PN WO9928482-A2.  
XX PD 10-JUN-1999.  
XX PF 24-NOV-1998; 98WO-US24528.  
XX PR 28-JUL-1998; 98US-0094331.  
XX PR 28-NOV-1997; 97US-0067315.  
XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.  
XX PT New hepatitis C virus covalent complexes  
XX PS Claim 6; Page 95-97; 21pp; English.  
XX CC The present invention describes a covalent hepatitis C virus (HCV)  
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
CC to the amino terminus of the HCV NS3 protease domain. The present  
CC sequence represents a specifically claimed example of the above  
CC complex. The covalent NS4A-NS3 complexes are useful for structural  
CC determination and determination of mode of binding of HCV inhibitors by  
CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
CC protease activity, the helicase activity and the ATPase activity of NS3.  
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
CC the non-covalent protease-peptide complexes previously available.  
XX SQ Sequence 665 AA;

Query Match 86.08; Score 874.5; DB 20; Length 665;  
Best Local Similarity 84.78; Pred. No. 4.2e-81;  
Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;  
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQCOETSGTGRDKNQVEGEVQIVSTATQTFLAT 61  
DB 22 GSVVIVGRILLSGSGSITAYSQOTRGLGCKKTSLTGRDKNQVEGEVQIVSTATQSFAT 81  
QY 62 SINGVLWTVYHGAGTRTIASPKGPVTOMYTNVDKLVGWQAPGSGSLTPTCTCGSSDLYL 121  
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVGMQAPPGARSULTPTCTCGSSDLYL 141  
QY 122 VTRHADVIPVRRRGDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVIFRAAVSTRGVAKA 181  
DB 142 VTRHADVIPVRRRGDSRGSLLSPRPVSYLGKSGGGLLCPSGHAGVIFRAAVCTRGVAKA 201  
QY 182 VDFIPVESLETTMRS 197  
DB 202 VDFVPVESMETTMRSP 217

RESULT 15  
AAAY24946  
ID AAY24946 standard; Protein: 665 AA.

XX AC AAY24946;  
XX DT 07-SEP-1999 (first entry)  
XX DE HCV NS4A-NS3 complex SEQ ID NO:17.  
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
KW hydrophobic domain; covalent complex; detection; inhibitor.  
XX OS Hepatitis C virus.  
OS Synthetic.

XX PN WO9928482-A2.  
XX PD 10-JUN-1999.  
XX PF 24-NOV-1998; 98WO-US24528.  
XX PR 28-JUL-1998; 98US-0094331.  
XX PR 28-NOV-1997; 97US-0067315.  
XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
XX WPI: 1999-385385/32.

```

XX New hepatitis C virus covalent complexes
XX
XX Claim 6; Page 97-99; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX
XX Sequence 665 AA:
SQ
    Query Match      86.0%; Score 874.5; DB 20; Length 665;
    Best Local Similarity 84.7%; Pred. No. 4.2e-81;
    Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEEGCOETISQTRGDKNQVEGEVQIVSTATQIFLAT 61
Db  |||||  |||  |||:||||  ||  ||  |||||  |||||:|||||
22 GSVVIVGRILLSGSITAYSQOTRGLLGC1KTSLTGRDKNQVEGEVQVYSTATQSFAT 81

QY 62 SINGVLWTVYHGAGTRTITASPKGPVTOMYTNVDKDLVQWQAPQGSRLTPTCGSSDLYL 121
Db  :|||  |||||  |||:|  |||||  |||||  |||||  |||||  |||||  |||||
82 CVNGVCWTVYHGAGSKTLAGPKGPITOMYTNVDQDLVQWQAPPGARSLTPTCGSSDLYL 141

QY 122 VTRHADVIPVRRRGDSRGSLSPRPISYIKSGGGLICPAGHAVGIFRAAVSTRGVAKA 181
Db  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
142 VTRHADVIPVRRRGDSRGSLSPRPVSYLKSGAGPLLCPSGHAVGIFRAAVCTRGVAKA 201

QY 182 VDFIPVESLETHRSP 197
Db  |||:||||:|||||
202 VDFVPVESMETMRSP 217
```

Search completed: August 30, 2003, 19:12:24  
Job time : 45.6227 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	854.5	84.0	3011	1	GNWVC3	genome polyprotein
2	853.5	83.9	3011	1	S40770	genome polyprotein
3	848.5	83.4	3011	1	GNWVCH	genome polyprotein
4	837.5	82.4	3010	1	GNWVTH	genome polyprotein
5	827.5	81.4	3010	1	A45573	genome polyprotein
6	823.5	81.0	3010	1	GNWVTC	genome polyprotein
7	823.5	81.0	3010	1	GNWVCJ	genome polyprotein
8	812.5	79.9	3010	1	S18030	genome polyprotein
9	743.5	73.1	3014	1	JC5620	genome polyprotein
10	675	66.4	3033	1	GNWVJ8	genome polyprotein
11	674	66.3	3033	1	JQ1303	genome polyprotein
12	249	24.5	3005	2	T08841	polyprotein - dour
13	243	23.9	2970	2	T08839	polyprotein - marm
14	92.5	9.1	590	2	B81104	nitrate/nitrite se
15	92.5	9.1	590	2	C81911	nitrate/nitrite se
16	85.5	8.4	209	2	H83144	probable aromatic
17	85.5	8.4	398	2	B71284	probable periplasm
18	83.5	8.2	1334	2	AB1775	hypothetical prote
19	83	8.2	452	2	I39383	angio-associated m
20	82.5	8.1	716	2	G83612	hypothetical prote
21	81	8.0	377	2	A75335	hypothetical prote
22	80	7.9	322	2	D87603	glycosyl transfera
23	80	7.9	477	2	E75392	hypothetical prote
24	80	7.9	915	2	F81196	transferrin-bindin
25	80	7.9	1615	2	JE0372	low density lipopr
26	79.5	7.8	514	2	AE2827	serine proteinase
27	79.5	7.8	2638	1	A42545	genome polyprotein
28	79	7.8	479	2	H70847	probable oxidoredu
29	79	7.8	3414	1	GNWVNE	genome polyprotein

```

Db      1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATSLNGVLVTVYHGAGTRTIIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
DB      1065 AAOTFLATCINGCVTVYHGAGTRTIIASPKGPVIQMYTNVDQDLVGVQAPQGSRLTPCT 1124
QY      114 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
DB      1125 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchausti, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein M #status predicted <EPM>
F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: hepatitis virus #status predicted <NS2>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <N5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <N5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match      83.9%; Score 853.5; DB 1: Length 3011;
Best Local Similarity 82.8%; Pred. No. 3.4e-69;
Matches 169; Conservative 8; Mismatches 18; Indels 9; Gaps 1;

QY      3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOETQOTGRDKNQVEGEVQIVST 53
DB      1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATSLNGVLVTVYHGAGTRTIIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
DB      1065 AAOTFLATCINGCVTVYHGAGTRTIIASPKGPVIQMYTNVDQDLVGVQAPQGSRLTPCT 1124
QY      114 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
DB      1125 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
S40770
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```

```

RESULT 3
GNMVCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
R:Inchausti, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MUID:92052256; PMID:1658800
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: major envelope protein M #status predicted <EPM>
F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>
F:1007-1615/Product: hepatitis virus #status predicted <NS2>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: DEXH motif
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <N5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <N5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match      83.4%; Score 848.5; DB 1: Length 3011;
Best Local Similarity 81.9%; Pred. No. 9.7e-69;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

QY      3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQCOETQOTGRDKNQVEGEVQIVST 53
DB      1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQOTRGLLGCIIITSLTRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATSLNGVLVTVYHGAGTRTIIASPKGPVTOMYTNVDKDLVGVQAPQGSRLTPCT 113
DB      1065 AAOTFLATCINGCVTVYHGAGTRTIIASPKGPVIQMYTNVDQDLVGVQAPQGSRLTPCT 1124
QY      114 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 173
DB      1125 CGSSDLVLYTRHADVIPVRRGRDGRSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 4
GNMVCH
genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992

```



F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 81.4%; Score 827.5; DB 1; Length 3010;  
Best Local Similarity 77.0%; Pred. No. 8e-67;  
Matches 157; Conservative 20; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KKGSVIVGRIN-----LSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVST 53  
Db 1005 RRGREILLGPADSIEGGWRLLAPITAYAQOTRGLGCIITSLTRDKNQVEGEVQIVST 1064

Qy 54 ATOTFLATSIINGVLWTVYHGAGTRTIASPGQVPTQMTYNDKDLVWGQAPGQSRSLTPCT 113  
Db 1065 ATQSFATCVNGVQWTVYHGAGSKTLACPKGPIQMTYNDQDLVWGHPGARGSLTPCT 1124

Qy 114 CGSSDLYLVTRHADVIPRRRGDSRGLSPRPISYLYKGGSGGPLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLYLVTRHADVIPRRRGDSRGLSPRPISYLYKGGSGGPLLCPAGHAGVIFRAAV 1184

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTGKVAKAVDFIPVESMETTRSP 1208

RESULT 6  
GNMVTG  
genome polypeptide - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstr)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C:Accession: A38465  
R:Takanizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from hu  
A:Reference number: A38465; MUID:91140698; PMID:1847440  
A:Accession: A38465  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <TAX>  
A:Cross-references: ENBL:M59335; NID:329770; PIDN:AAA72945.1; PID:G329771  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <ME>  
F:390-728/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 81.0%; Score 823.5; DB 1; Length 3010;  
Best Local Similarity 76.5%; Pred. No. 1.9e-66;  
Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KKGSVIVGRIN-----LSGDTAYAQOTRGEQCOETSGTRDKNQVEGEVQIVST 53  
Db 1005 RRGREILLGPADSIEGGWRLLAPITAYAQOTRGLGCIITSLTRDKNQVEGEVQIVST 1064

Qy 54 ATOTFLATSIINGVLWTVYHGAGTRTIASPGQVPTQMTYNDKDLVWGQAPGQSRSLTPCT 113  
Db 1065 ATQSFATCVNGVQWTVYHGAGSKTLAAPKGPITQMTYNDQDLVWGHPGARGSLTPCT 1124

Qy 114 CGSSDLYLVTRHADVIPRRRGDSRGLSPRPISYLYKGGSGGPLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLYLVTRHADVIPRRRGDSRGLSPRPISYLYKGGSGGPLLCPAGHAGVIFRAAV 1184

A:Variety: isolate JKI  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C:Accession: S18030; S33570; A48332; S18029  
R:Honda, M.; Kaneo, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single pa  
A:Reference number: S18028  
A:Accession: S18030  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <HOW>  
A:Cross-references: EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
A:Experimental source: isolate JKI from an individual  
R:Honda, M.; Kaneo, S.; Unoura, K.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolat  
A:Reference number: A48332; MUID:93119270; PMID:9380322  
A:Accession: S33570  
A:Molecule type: genomic RNA  
A:Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
as Trp, and TTC for residue 771 as Ser  
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-131/Product: envelope protein M #status predicted <EPM>  
F:192-329/Product: major envelope protein E #status predicted <ME>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus nucleotide binding motif A (P-loop)  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate  
Query Match 79.9%; Score 812.5; DB 1; Length 3010;  
Best Local Similarity 76.0%; Pred. No. 1.9e-65;  
Matches 155; Conservative 20; Mismatches 20; Indels 9; Gaps 1;  
QY 3 KKGSVVIVGRIN-----LSGGDTAYAQOTRGEEGCOETSGTGRDKNQVEGEQIVST 53  
DB 1005 RRGREILLGPDGFRGQWRLLAPITAYSQOTRGLFGCIVTSLTGRDKNQVEGEQIVST 1064  
QY 54 ATCTFLATISGVLYTVYHICAGTRTTTASPKGPYQMTYVNDKDLVQWAPQGSRLTPCT 113  
DB 1065 ATQSFPLATCVNGVMTYVHGAGSKTLGAPKPGINQMTYVNDQDLVQWAPSGAASLTPTCT 1124  
QY 114 CGSSDLVYTRHADVIPVRRRGRSGLLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 173  
DB 1125 YGSSDLVYTRHADVIPVRRRGRSGLLSPRPVSYLKSGSGGPLLCPSGHANGIFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
DB 1185 CTRGVAKAVDFIPVESMETTMRSP 1208  
RESULT 9  
JC5620  
genome polyprotein - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus nucleotide binding motif A; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: JC5620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina  
A:Reference number: JC5620; MUID:97366593; PMN:9223423

```
Query Match      66.3%; Score 674; DB 1; Length 3033;
Best Local Similarity 68.7%; Pred. No. 8.4e-53;
Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;
```



A: Molecule type: DNA  
A: Residues: 1-590 <PAR>  
A: Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84658.1; PID:g738007  
A: Experimental source: serogroup A, strain Z2491  
C: Genetics:  
A: Gene: NMA1418  
C: Superfamily: nitrate/nitrite sensor protein narX  
C: Keywords: autophosphorylation; phosphohistidine; phosphohistidine; phosphotransferase; s  
F:395/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 9.18; Score 92.5; DB 2; Length 590;  
Best Local Similarity 21.3%, Pred. No. 1.5;  
Matches 46; Conservative 26; Mismatches 79; Indels 65; Gaps 6;

QY 28 BEGQETQGRDRKNOVEGEVQIVSTATQTFATSIINGVLMTVYHGAGTRTIASPKGPVT 87  
DB 213 EGGTFEFKQVGRCFNQMGRLKILYDDLEGQVAEQ-----TRSLKQONQNL 259  
QY 88 QMYTNVDKDLVGMQAPQ-----GSRSLTPCTCGSSDLYLVTRHAD----- 127  
DB 260 LLY-OTTRDLHQSYPQAAEHFLNRLPAVGADSGRVCLDGGSDVYVSIHHADCGTAAS 318  
QY 128 -----VIVRRGDSRGSLLSPRISYLYKSGSGPLLCPAGHAVGIFRAAYSTR--- 176  
DB 319 DLGKYHEEIPLEYQNETPLGRLLISFPNGISLDEDDRIQLTLGRQLGVSLAGAKQEEK 378  
QY 177 -----GVAKAVDF--IPVESLET 192  
DB 379 RLLAVLQERNLIAQGLHDSIAQALTFLNLQVOMLET 414

Search completed: August 30, 2003, 19:20:30  
Job time : 17.2134 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	854.5	84.0	3011	1	POLG_HCV1		P26664 h genome po
2	848.5	83.4	3011	1	POLG_HCV1		P27958 h genome po
3	837.5	82.4	3010	1	POLG_HCVTW		P29846 h genome po
4	827.5	81.4	3010	1	POLG_HCVTW		Q00269 h genome po
5	823.5	81.0	3010	1	POLG_HCVBK		P26663 h genome po
6	823.5	81.0	3010	1	POLG_HCVJA		P26662 h genome po
7	675	66.4	3033	1	POLG_HCVJ8		P26661 h genome po
8	674	66.3	3033	1	POLG_HCVJ6		P26660 h genome po
9	87	8.6	321	1	HHOA_ARATH		Q9sel7 arabidopsis
10	85.5	8.4	409	1	PAAD_PSEAE		Q9hx08 pseudomonas
11	83.5	8.2	437	1	DEGL_ARATH		Q22609 arabidopsis
12	83	8.2	452	1	AAMP_HUMAN		Q13685 homo sapien
13	79.5	7.8	3414	1	POLG_LANVT		P29837 t genome po
14	79	7.8	3414	1	POLG_TBENV		P14336 t genome po
15	78.5	7.7	706	1	TREE_HORSE		P27425 equus caball
16	78.5	7.7	764	1	ICCR_DROME		Q08180 drosophila
17	78	7.7	911	1	TB11_NEIMB		Q09056 neisseria m
18	78	7.7	3412	1	POLG_TBENV		P07720 t genome po
19	77.5	7.6	263	1	GRAK_MOUSE		Q35205 mus musculu
20	77	7.6	594	1	NIR_SPIOL		P05314 spinacia ol
21	77	7.6	1705	1	PTPO_MOUSE		P70289 mus musculu
22	77	7.6	3414	1	POLG_TBENV		Q01299 t genome po
23	76.5	7.5	323	1	VPRT_SMRVH		P21407 squirrel mo
24	76.5	7.5	333	1	MOSA_RHIME		Q07607 rhizobium m
25	76.5	7.5	452	1	MLTD_ECOLI		P23931 escherichia
26	76	7.5	3411	1	POLG_YEFV1		P03314 y genome po
27	76	7.5	3411	1	POLG_YEFV2		P19901 y genome po
28	75.5	7.4	485	1	Y136_TREPA		O83172 treponema p
29	75.5	7.4	2269	1	WDR9_HUMAN		Q9nsi6 homo sapien
30	75	7.4	467	1	NX1B_BOVIN		Q28142 bos taurus
31	75	7.4	973	1	VP18_HUMAN		Q9p253 homo sapien
32	75	7.4	1165	1	POLG_GALY		P21414 gibbon ape
33	74.5	7.3	1248	1	GRAD_MOUSE		P11033 mus musculu

InterPro: IPR002521; HCV\_core.  
 InterPro: IPR002519; HCV\_env.  
 InterPro: IPR002531; HCV\_NS1.  
 InterPro: IPR002538; HCV\_NS2.  
 InterPro: IPR004109; HCV\_NS3.  
 InterPro: IPR000745; HCV\_NS4a.  
 InterPro: IPR001490; HCV\_NS4b.  
 InterPro: IPR002868; HCV\_NS5a.  
 InterPro: IPR002166; HCV\_RdRp.  
 InterPro: IPR001650; Helicase\_C.  
 InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 InterPro: IPR007094; RNA\_pol\_PSVir.  
 Pfam: PF01543; HCV\_capsid; 1.  
 Pfam: PF01542; HCV\_core; 1.  
 Pfam: PF01539; HCV\_env; 1.  
 Pfam: PF01560; HCV\_NS1; 1.  
 Pfam: PF01538; HCV\_NS2; 1.  
 Pfam: PF02907; HCV\_NS3; 1.  
 Pfam: PF01006; HCV\_NS4a; 1.  
 Pfam: PF01001; HCV\_NS4b; 1.  
 Pfam: PF01506; HCV\_NS5a; 1.  
 Pfam: PF00271; helicase\_C; 1.  
 Pfam: PF00998; Viral\_RdRp; 1.  
 ProDom: PD186062; HCV\_NS1; 1.  
 SMART: SM00487; DEXDc; 1.  
 PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 3D-structure.  
 INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 CELLULAR AMINOPEPTIDASE.  
 CHAIN 1 115  
 CHAIN 116 191  
 CHAIN 192 383  
 CHAIN 384 729  
 CHAIN 730 1006  
 CHAIN 1007 1615  
 CHAIN 1616 1862  
 CHAIN 1863 2013  
 CHAIN 2014 3011  
 CHAIN 3012 369  
 CHAIN 369 1083  
 CHAIN 1083 1107  
 CHAIN 1107 1165  
 CHAIN 1165 1237  
 CHAIN 1237 1319  
 CHAIN 1319 196  
 CHAIN 196 209  
 CHAIN 209 234  
 CHAIN 234 305  
 CHAIN 305 417  
 CHAIN 417 423  
 CHAIN 423 430  
 CHAIN 430 448  
 CHAIN 448 476  
 CHAIN 476 532  
 CHAIN 532 540  
 CHAIN 540 556  
 CHAIN 556 576  
 CHAIN 576 623  
 CHAIN 623 645  
 CHAIN 645 2041  
 CHAIN 2041 2077  
 CHAIN 2077 2240  
 CHAIN 2240 2364  
 CHAIN 2364 2789  
 CHAIN 2789 3011  
 CHAIN 3011 AA; 327197 MW; 65F8C9447FCB5AF9 CRC64;  
 Query Match 84.08; Score 854.5; DB 1; Length 3011;  
 Best Local Similarity 82.88; Pred. No. 2.1e-71;  
 Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;  
 3 KGSWIVGRIN-----LSGDTAYAAQTRGECGQETSQTRDKNQVGEVIVST 53

Db 1005 RRGREILGADGMVSKGWRLLAPITAYAAQTRGCGLLGCIITSLTRDKNQVGEVIVST 1064  
 Qy 54 ATQTFLATSLNGVLVYHGAGTRTIAIPRGVPTQMTNTVDKDLVQWAPQGSRLTPTCT 113  
 Db 1065 AAOTFLATSLNGVCTVYHGAGTRTIAIPRGVPTQMTNTVDKDLVQWAPQGSRLTPTCT 1124  
 Qy 114 CGSSDLYLVTRHADVIPIVRRGDSRGLSPRISYLYKSGSGGGLLCPAGHAGVIFRAAV 173  
 Db 1125 CGSSDLYLVTRHADVIPIVRRGDSRGLSPRISYLYKSGSGGGLLCPAGHAGVIFRAAV 1184  
 Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RA "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=9733322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RA "Structure of the hepatitis C virus RNA helicase domain.";  
 RA Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RA "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1',  
 CC [RNA]{N}.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC [RNA]{N}.  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.



```

CC CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@sib-sib.ch).
CC CC -----
CC CC EMBL: M67463; AAA4534.1; .
DR DR PIR: A36814; GNMVCH.
DR DR PDB: 1HEI; 25-NOV-98.
DR DR PDB: 1AIV; 16-FEB-99.
DR DR PDB: 1A1R; 17-JUN-98.
DR DR MEROPS: S29.001; .
DR DR MEROPS: U39.001; .
DR DR TRANSFAC: T04155; .
DR DR InterPro: IPR001410; DEAD.
DR DR InterPro: IPR002522; HCV capsid.
DR DR InterPro: IPR002521; HCV core.
DR DR InterPro: IPR002519; HCV_env.
DR DR InterPro: IPR002531; HCV_NS1.
DR DR InterPro: IPR002518; HCV_NS2.
DR DR InterPro: IPR004109; HCV_NS3.
DR DR InterPro: IPR000745; HCV_NS4a.
DR DR InterPro: IPR001490; HCV_NS4b.
DR DR InterPro: IPR002868; HCV_NS5a.
DR DR InterPro: IPR002166; HCV_RdRp.
DR DR InterPro: IPR001650; Helicase_C.
DR DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR DR InterPro: IPR007094; RNA_pol_PSVir.
DR DR Pfam: PF01543; HCV_capsid; 1.
DR DR Pfam: PF01542; HCV_core; 1.
DR DR Pfam: PF01539; HCV_env; 1.
DR DR Pfam: PF01560; HCV_NS1; 1.
DR DR Pfam: PF01538; HCV_NS2; 1.
DR DR Pfam: PF02907; HCV_NS3; 1.
DR DR Pfam: PF01006; HCV_NS4a; 1.
DR DR Pfam: PF01001; HCV_NS4b; 1.
DR DR Pfam: PF01506; HCV_NS5a; 1.
DR DR Pfam: PF00271; helicase_C; 1.
DR DR Pfam: PF00998; Viral_RdRp; 1.
DR DR ProDom: PD186062; HCV_NS1; 1.
DR DR SMART: SM00487; DEXDC; 1.
KW KW Polypeptide: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
FW FW 3D-structure.
FW FW INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FW FW CHAIN 1 191 CELLULAR AMINOPEPTIDASE.
FW FW CHAIN 192 383 CAPSID PROTEIN C.
FW FW CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
FW FW CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
FW FW CHAIN 810 1026 PROTEIN P7.
FW FW CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
FW FW CHAIN 1658 1711 PROTEASE/HELICASE NS3.
FW FW CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4a.
FW FW CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4b.
FW FW CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5a.
FW FW CHAIN 3011 369 NONSTRUCTURAL PROTEIN NS5b.
FW FW TRANSMEM 347 369 POTENTIAL.
FW FW ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FW FW ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FW FW ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FW FW NP_BIND 1230 1237 SITE (POTENTIAL).
FW FW SITE 1316 1319 DECH BOX.
FW FW CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FW FW CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FW FW CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1224 1226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1232 1233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1236 1238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1239 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1247 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1251 1255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1258 1271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1272 1272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1277 1280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1281 1282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1283 1285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1291 1295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1296 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1302 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1312 1316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1317 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1323 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1336 1340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1343 1347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1352 1353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1360 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1362 1366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1368 1368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1373 1375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1376 1377 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1378 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1382 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1389 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1397 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1410 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1414 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1419 1420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1432 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1438 1439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1450 1453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1456 1463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1471 1478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1480 1480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1481 1488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1489 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1507 1507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1511 1511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1514 1527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1532 1544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1555 1564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1570 1578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1579 1580 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1584 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1598 1598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1606 1611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 1614 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1622 1623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1627 1627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1635 1636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX 1640 1652 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

```

Query Match

83.4%; Score 848.5; DB 1; Length 3011;

Best Local Similarity

81.9%; Pred, No. 7.5e-71;



Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;  
QY 3 KKGWVIVGRIN-----LSGDYAYAQQTGRGEGCOETSGTGRKKNQVEGEVQIVST 53  
DB 1005 RRQGEILLGADGWSKGNWFLAPITAYAAQTGRGLGCIITSLTGRKKNQVEGEVQIVST 1064  
QY 54 ATQTFLATISINGVLTWYHAGAGTRTTASPKGPVQTYTWNVDLVGWAQPOGSRSLTPCT 113  
DB 1065 ATOTFLATCINGVLTWYHAGAGTRTTASPKGPVQTYTWNVDLVGWAQPOGSRSLTPCT 1124  
QY 114 CGSSDLYLVTRHADVTPVRGRDGRSGLSPRISYLVKSSGGLLCPAGHAGVIFRAAV 173  
DB 1125 CGSSDLYLVTRHADVTPVRGRDGRSGLSPRISYLVKSSGGLLCPAGHAGVIFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208  
RESULT 3  
POLG\_HCVTM STANDARD; PRT: 3010 AA.  
AC P29846;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Taiwan) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31645;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=92230206; PubMed=1314449;  
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
RA "The Taiwanese hepatitis C virus genome: sequence determination and  
RA mapping the 5' termini of viral genomic and antigenomic RNA.";  
RL Virology 188:102-113(1992).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: M84754; ; NOT\_ANNOTATED\_CDS.  
DR PIR: A40244; GNMVTV  
DR PDB: 1N64; 25-FEB-03.  
DR PDB: 1NS3; 08-APR-98.  
DR MEROPS: S29.001; -;  
DR MEROPS: U39.001; -;  
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRP.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; helicase\_C; 1.  
DR Pfam: PF00998; Viral\_RdRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART; SMO0487; DEXDC; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure.  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 233 233  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2529 2529  
FT CARBOHYD 2788 2788  
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;  
Query Match 82.4%; Score 837.5; DB 1; Length 3010;  
Best Local Similarity 78.4%; Pred. No. 8e-70;  
Matches 160; Conservative 18; Mismatches 17; Indels 9; Gaps 1;





FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT HELIX 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA; F8422D5ECCDFD9C CRC64;

Query Match 81.0%; Score 823.5; DB 1; Length 3010;  
 Best Local Similarity 76.5%; Pred. No. 1.6e-68;  
 Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;  
 QY 3 KKGSVVIVGRIN-----LSGDTAYAOOTRGEQCETSGTGRDKNOVGEVQIVST 53  
 DB 1005 RRGKILLGPADSLRGLRLLAPITAYTSQOTRGLGCIITSLTGRDKNOVGEVQIVST 1064  
 QY 54 ATOTFLATISINGVLTVYHAGTRTIASPKGPVOMYTNVDKDLVGNQAPGSGSLPPT 113  
 DB 1065 ATQSEFLATCVNVCVTYVHGAGSKTLAAPKGPITOMYTNVDQDLVGNPKPGARSLLPCT 1124  
 QY 114 CGSSDLYLVTRHADYIPVRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAV 173  
 DB 1125 CGSSDLYLVTRHADYIPVRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184  
 QY 174 STRGAKAVDFIPVESLETTMRSP 197  
 DB 1185 CTRGAKAVDFIPVESMETMRSP 1208

## RESULT 6

POLG\_HCVJA STANDARD; PRT; 3010 AA.  
 AC P26662;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Envelope glycoprotein (Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV).  
 OS Hepatitis C virus (isolate Japanese) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088550; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from  
 RT Japanese patients with non-A, non-B hepatitis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=91192160; PubMed=1849488;  
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraio K.,  
 RA Ohkoshi S., Shimotohno K.;  
 RT "Molecular structure of the Japanese hepatitis C viral genome";  
 RL FEBS Lett. 280:325-328(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC {RNA}(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D90208; BAAL4233.1;  
 CC PIR: A39253; GNMVCJ.  
 CC HSP: P26663; LJXP.  
 CC MEROPS: S29.001;  
 CC MEROPS: U39.001;  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002521; HCV\_capsid.  
 CC InterPro: IPR002522; HCV\_core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC InterPro: IPR002531; HCV\_NS1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR004109; HCV\_NS3.  
 CC InterPro: IPR000745; HCV\_NS4a.  
 CC InterPro: IPR001490; HCV\_NS4b.  
 CC InterPro: IPR002868; HCV\_NS5a.  
 CC InterPro: IPR002166; HCV\_RdRp.

DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR Pfam: PF0186062; HCV\_NS3; 1.  
 DR Pfam: PF0186062; HCV\_NS3; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 DR INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT CHAIN 3010 369  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2788 2788  
 FT SEQUENCE 3010 AA; 327017 MW; AA993794F460B185 CRC64;  
 Query Match 81.08; Score 823.5; DB 1; Length 3010;  
 Best Local Similarity 75.58; Pred. No. 1.6e-68;  
 Matches 154; Conservativity 23; Mismatches 18; Indels 9; Gaps 1;  
 QY 3 KKGSVVIVGRINLSGD-----TAYAQTRGEGCQTSOTGRKNQVEGVIVST 53  
 DB 1005 RRGKEILLGPADSFGEQGWRLAPITAYSQQTGRLGCIITSLTGRKNQVDEGVVLST 1064  
 QY 54 ATQFLATISGVLTWYTHGAGTITIASPKGPTQMTYNDKDLVQWAPQGSRLTPTCT 113  
 DB 1065 ATQSFALTCVNGVCTWYVHGAGSKTLAGPKPIQMTYNDQDLVGPAPPGASMTPTCT 1124  
 QY 114 CGSSDLYLVTRHADVIVRRRGDSRGSLSPRISYLGSGGGPLLCAGHAYGIFRAAV 173  
 DB 1125 CGSSDLYLVTRHADVIVRRRGDSRGSLSPRISYLGSGGGPLLCPSGVHVGIFRAAV 1184  
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197

Db 1185 CTGVAKAVDFIPVESKETTMRSP 1208  
 RESULT 7  
 ID POLG\_HCVJ8 STANDARD; PRT; 3033 AA.  
 AC P26661;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein  
 DE NS4b (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J8) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirin.  
 CC NCBI\_TaxID=11115;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=92230232; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda S., Mishiro S.;  
 RT \*Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes\*;  
 RT Virology 188:331-341(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC EMBL: D10988; BAA01761.1; -;  
 DR PIR: A40250; GNMVJ8.  
 DR HSP: P27958; 1HEI.  
 DR MEROPS: S29.001; -;  
 DR MEROPS: U39.001; -;  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_RdRp.  
 DR InterPro: IPR002166; HCV\_NS5a.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.

DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00998; Viral\_RdRp; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT ACT\_SITE 1234 1241  
 FT BIND 1320 1323  
 FT SITE 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 299 299  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2359 2359  
 FT CARBOHYD 2811 2811  
 FT SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;  
 Query Match 66.4%; Score 675; DB 1; Length 3033;  
 Best Local Similarity 69.8%; Pred. No. 1.3e-54;  
 Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;  
 QY 19 TAYAAQTGEECCQTSOTGRDKNOVEGVIVSTATQTFLATSGINGVLTWYHGAGTRT 78  
 DB 1034 TAYTOOTRGLLGAIVYSLTGRDKNEQAGQVQLSVTQTGLTSGISGLVLTWYHGAGNKT 1093  
 QY 79 IASPGKPTQMYTNDKLVGKQVQPGSRSITPCTCGSSDLVLTNRADVIPVRRGDSR 138  
 DB 1094 LAGPKGPTQMYTSEGDLVGVSPGPKSLDPTCGNVDLYLTNRADVIPVRKDDRR 1153  
 QY 139 GSLLSPRTSYLKGSGGGLLCPAGHAVGIFRAAVSTRGVAKAVDFIPVESLETHMRSP 197  
 DB 1154 GALLSPRLSTLKGSGGGLVLCRSHAVGLFRAAVCAARGVAKSIDFIPVESLDVATRTP 1212  
 RESULT 8  
 ID POLG\_HCVJ6 STANDARD; PRT: 3033 AA.  
 AC P26660;

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticvirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus [isolate HC-J6] (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92044440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. gen. Virol. 72:2697-2704(1991).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MENA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S39.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: D00944; BAA00792.1; -.  
 CC PIR: JQ1303; JQ1303.  
 CC HSSP: P27958; 1HEI.  
 CC MEROPS: S29.001; -.  
 CC MEROPS: U39.001; -.  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV\_capsid.  
 CC InterPro: IPR002521; HCV\_core.  
 CC InterPro: IPR002519; HCV\_env.  
 CC InterPro: IPR002531; HCV\_NS1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR004109; HCV\_NS3.  
 CC InterPro: IPR000745; HCV\_NS4a.  
 CC InterPro: IPR001490; HCV\_NS4b.  
 CC InterPro: IPR002868; HCV\_NS5a.  
 CC InterPro: IPR002166; HCV\_RdRp.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro: IPR007094; RNA\_pol\_PSwir.  
 CC Pfam: PF01543; HCV\_capsid; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01539; HCV\_env; 1.  
 CC Pfam: PF01560; HCV\_NS1; 1.  
 CC Pfam: PF01538; HCV\_NS2; 1.  
 CC Pfam: PF02907; HCV\_NS3; 1.  
 CC Pfam: PF01006; HCV\_NS4a; 1.  
 CC Pfam: PF01001; HCV\_NS4b; 1.  
 CC Pfam: PF01506; HCV\_NS5a; 1.









RESULT 12

Db 296 GSWDCQAKLVSAATTGKVVGVFRPETVASQPSLGRGSEESNSVESL 341

## RESULT 13

POID\_LANVT STANDARD: PRT: 3414 AA.

AC DT 01-APR-1993 (Rel. 25, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [contains: Capsid protein C (Core protein);

DE Envelope protein PM; Matrix protein (Envelope protein M); Major

DE Envelope protein E; Nonstructural protein NS1; Nonstructural protein

DE NS2A; Nonstructural protein NS2B; Helicase/protease (EC 3.4.21.98)

DE (NS3); Nonstructural protein NS4A; Nonstructural protein NS4B; RNA-

DE directed RNA polymerase (EC 2.7.7.48) (NS5)].

OS Langkat virus (strain TP21).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus.

OC NCBI\_TaxID=31638;

RN [1]

RP SEQUENCE OF 1-776 FROM N.A.

RX MEDLINE=92074260; PubMed=1720591;

RA Mandl C.W., Iacono-Connors L., Wallner G., Holzmann H., Kunz C.,

RA Heinz F.X.;

RT \*Sequence of the genes encoding the structural proteins of the low-

RT virulence tick-borne flaviviruses Langat TP21 and Yelantsev.\*;

RN Virology 185:891-895(1991).

RN [2]

RP SEQUENCE OF 777-3414 FROM N.A.

RX MEDLINE=92263794; PubMed=1316884;

RA Iacono-Connors L.C., Schmaljohn C.S.;

RT \*Cloning and sequence analysis of the genes encoding the

RT nonstructural proteins of Langat virus and comparative analysis with

RT other flaviviruses.\*;

RN Virology 188:875-880(1992).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +

CC (RNA)(N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M73835; AAA02740.1; ALT\_TERM.

DR EMBL: S35365; AAB22165.1; .

DR PIR: A42545; A42545.

DR HSSP: P14336; LSVB.

DR MEROPS: S07.001; .

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001122; Flavi\_capsidC.

DR InterPro: IPR000336; Flavi\_glycoprote.

DR InterPro: IPR001850; Flavi\_helicase.

DR InterPro: IPR000069; Flavi\_M.

DR InterPro: IPR001157; Flavi\_NS1.

DR InterPro: IPR000752; Flavi\_NS2A.

DR InterPro: IPR000487; Flavi\_NS2B.

DR InterPro: IPR000404; Flavi\_NS4A.

DR InterPro: IPR001528; Flavi\_NS4B.

DR InterPro: IPR000208; Flavi\_NS5.

DR InterPro: IPR002535; Flavi\_propep.

DR InterPro: IPR002877; FtsJ.

DR InterPro: IPR001650; Helicase\_C.

DR InterPro: IPR007095; RNA\_pol\_DS\_PS.

DR InterPro: IPR007094; RNA\_pol\_PSVir.

DR Pfam: PF01003; Flavi\_capsid; 1.

DR Pfam: PF02832; Flavi\_glycop\_C; 1.

DR Pfam: PF00869; Flavi\_glycoprot; 1.

DR Pfam: PF00949; Flavi\_helicase; 1.

DR Pfam: PF01004; Flavi\_M; 1.

DR Pfam: PF00948; Flavi\_NS1; 1.

DR Pfam: PF01005; Flavi\_NS2A; 1.

DR Pfam: PF01002; Flavi\_NS2B; 1.

DR Pfam: PF01350; Flavi\_NS4A; 1.

DR Pfam: PF01349; Flavi\_NS4B; 1.

DR Pfam: PF00972; Flavi\_NS5; 1.

DR Pfam: PF01570; Flavi\_propep; 1.

DR Pfam: PF01728; FtsJ; 1.

DR Pfam: PF00271; helicase\_C; 1.

DR ProDom: PD001556; Flavi\_glycoprote; 1.

DR ProDom: PD001496; Flavi\_NS1; 1.

DR SMART: SM00487; DEXDc; 1.

DR SMART: SM00490; HELICc; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;

KW ATP-binding; Transmembrane; Nonstructural protein.

FT INIT\_MET 1 1

FT CHAIN 1 112

FT CHAIN 113 205

FT CHAIN 206 280

FT CHAIN 281 776

FT CHAIN 777 1128

FT CHAIN 1129 1358

FT CHAIN 1359 1489

FT CHAIN 1490 2110

FT CHAIN 2111 2259

FT CHAIN 2260 2511

FT CHAIN 2512 3414

FT NP\_BIND 1688 1695

FT SITE 1779 1782

FT TRANSMEM 103 119

FT TRANSMEM 262 278

FT TRANSMEM 728 744

FT TRANSMEM 758 774

FT DISULFID 283 310

FT DISULFID 340 396

FT DISULFID 354 385

FT DISULFID 372 401

FT DISULFID 466 570

FT DISULFID 587 618

FT CARBOHYD 144 144

FT CARBOHYD 434 434

SO SEQUENCE 3414 AA; 378017 MW; 59CB7E95DD70D82E CRC64;

Query Match 7.8%; Score 79.5; DB 1; Length 3414;

Best Local Similarity 22.7%; Pred. No. 69;

Matches 39; Conservative 21; Mismatches 67; Indels 45; Gaps 8;

QY 30 GCQETSQTGRDNQVEGEVQIVSTA-----TQTFLATSLNGVLWTVYH---GACRTRTAS 81

Db 1496 GCSEGRSDSRPLDVNRGVRYRTYPTGLLWQGRQIGVGYGAKGVLTMMHVTGCAALLVDGV 1555

QY 82 PKGPVTQMTYNDKDLV-----GMOA-----PGGSRSLTPTCGSSDLYLVT 123

Db 1556 AVCP---YNAVREDVVCYGGANSLESRWGRGTQVQVHAFPPG-RAHETHQCPQGLIL-- 1609

QY 124 RHADVIPVRRRGDSRGLSPRPISYKSGSGGLCPAGHAGVIFRAAVST 175

Db 1610 -----ENGRKMGAI----PIDLAKTSGSPIMNSOGVEVGYGLNGLKT 1648

RESULT 14

POLG\_TBEVN STANDARD: PRT: 3414 AA.

ID POLG\_TBEVN



ID	TRFE_HORSE	STANDARD;	PRT;	706 AA.
AC	P27425;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Serotransferrin precursor (transferrin) (Siderophilin) (Beta-1-metal binding globulin).			
DE	TF.			
GN	Equus caballus (Horse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.			
OC	NCBI_TaxID=9796;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=93277958; PubMed=8504171;			
RX	Carpenier M.A.; Broad T.E.;			
RT	*The cDNA sequence of horse transferrin.*;			
RL	Biochim. Biophys. Acta 1173:230-232(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=Extraembryonic tissue;			
RC	McDowell K.J.; Adams M.H.; Baker C.B.;			
RC	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING OF AN ANION, USUALLY BICARBONATE. IT IS RESPONSIBLE FOR THE TRANSPORT OF IRON FROM SITES OF ABSORPTION AND HEME DEGRADATION TO THOSE OF STORAGE AND UTILIZATION. SERUM TRANSFERRIN MAY ALSO HAVE A FURTHER ROLE IN STIMULATING CELL PROLIFERATION.			
CC	-!- SUBUNIT: MONOMER.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- TISSUE SPECIFICITY: Expressed in liver; secreted in plasma.			
CC	-!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.			
CC	-!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
EMBL	M69020; AAA30958.1; -			
DR	EMBL; U21127; AAA633684.1; -			
DR	PIR; S33761; S33761.			
DR	HSP; P02787; 1A8E.			
DR	InterPro: IPR001156; Transferrin.			
DR	Pfam: PF00405; transferrin; 2.			
DR	PRINTS; PR00422; TRANSFERRIN.			
DR	SMART; SM00094; TR_FER; 2.			
DR	PROSITE; PS00205; TRANSFERRIN_1; 2.			
DR	PROSITE; PS00206; TRANSFERRIN_2; 2.			
DR	PROSITE; PS00207; TRANSFERRIN_3; 2.			
DR	Transprot; Iron transport; Glycoprotein; Metal-binding; Repeat;			
KW	Signal.			
FT	SIGNAL. 1 19 BY SIMILARITY.			
FT	CHAIN 20 706 SEROTRANSFERRIN.			
FT	REPEAT 20 357 1.			
FT	REPEAT 358 706 2.			
FT	DISULFID 26 64 BY SIMILARITY.			
FT	DISULFID 36 55 BY SIMILARITY.			
FT	DISULFID 134 215 BY SIMILARITY.			
FT	DISULFID 174 190 BY SIMILARITY.			
FT	DISULFID 177 198 BY SIMILARITY.			
FT	DISULFID 187 200 BY SIMILARITY.			
FT	DISULFID 248 262 BY SIMILARITY.			
FT	DISULFID 360 623 BY SIMILARITY.			
FT	DISULFID 366 398 BY SIMILARITY.			
FT	DISULFID 376 389 BY SIMILARITY.			
FT	DISULFID 423 701 BY SIMILARITY.			
FT	DISULFID 441 664 BY SIMILARITY.			
FT	DISULFID 474 550 BY SIMILARITY.			

```
T DISULFID 498 692 BY SIMILARITY.
T DISULFID 508 522 BY SIMILARITY.
T DISULFID 519 533 BY SIMILARITY.
T DISULFID 590 604 BY SIMILARITY.
T DISULFID 642 647 BY SIMILARITY.
T METAL 79 79 IRON 1 (BY SIMILARITY).
T METAL 111 111 IRON 1 (BY SIMILARITY).
T METAL 209 209 IRON 1 (BY SIMILARITY).
T METAL 270 270 IRON 1 (BY SIMILARITY).
T METAL 413 413 IRON 2 (BY SIMILARITY).
T METAL 449 449 IRON 2 (BY SIMILARITY).
T METAL 544 544 IRON 2 (BY SIMILARITY).
T METAL 612 612 IRON 2 (BY SIMILARITY).
T BINDING 140 140 ANION (POTENTIAL).
T BINDING 480 480 ANION (POTENTIAL).
T CARBOHYD 515 515 N-LINKED (GLCNAC... ) (POTENTIAL).
T Q SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;

Query Match 7.7%; Score 78.5; DB 1; Length 706;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 48; Conservative 22; Mismatches 81; Indels 51; Gaps 11;

y 5 GSVIVIGRIMLSGDTAYAQTRGEGCGQETSQTRKKNQVEGEVQIVSTATQTFIATSIN 64
D 415 GFYIAGKCLVPVLAENYETRSACVDTPEGYH-----AVAVKSSSDPDLT---- 464
y 65 GVLWTVYHGAGTRTIAAPKGPVTOMVTNVDKLVGWAQPSRSLTPCTCGSSDLVLVTR 124
b 465 ---WN-----SLKCK-KSCHTGVDR-TAGWNIPMGL-----LYSEIK 496
y 125 HADVIPRRGDSRGLLSRPPISYLKGSGGP-LIC-PAGHA-----VGIPRAAVSTRG 177
b 497 HCEFDKFFREGCAPGYRRNSTLCNLICGSASGPGRECEPNNHERRYGYTGAFRCLEKGD 556
y 178 VAKAVDFIPVESLE--TTMRSP 197
b 557 VA----FVKHQTVEQNTDGRNP 574
```

Search completed: August 30, 2003, 19:13:48  
Job time : 11.7567 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 : Search time 37.5921 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriaph.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	872.5	85.8	4040	12	Q91FH8	Q91FH8 mucosal dis
2	858.5	84.4	3011	12	O36579	O36579 hepatitis c
3	854.5	84.0	2436	12	O81756	O81756 hepatitis c
4	854.5	84.0	3011	12	Q91FES	Q91FES hepatitis c
5	854.5	84.0	3011	12	Q91ELS8	Q91ELS8 hepatitis c
6	853.5	83.9	3011	12	O33463	O33463 hepatitis c
7	851.5	83.7	3011	12	O36608	O36608 hepatitis c
8	851.5	83.7	3015	12	O9PWX5	O9PWX5 hepatitis c
9	851.5	83.7	3015	12	Q9PWU9	Q9PWU9 hepatitis c
10	849	83.5	181	12	Q91RR8	Q91RR8 hepatitis c
11	849	83.5	181	12	Q91RT5	Q91RT5 hepatitis c
12	847	83.3	181	12	Q91RR5	Q91RR5 hepatitis c
13	847	83.3	181	12	Q91RR2	Q91RR2 hepatitis c
14	847	83.3	181	12	Q91RT9	Q91RT9 hepatitis c
15	846	83.2	181	12	Q91RR3	Q91RR3 hepatitis c
16	846	83.2	181	12	Q91RR4	Q91RR4 hepatitis c

# ALIGNMENTS

## RESULT 1

ID	Q91FH8	PRELIMINARY;	PRT: 4040 AA.
AC	Q91FH8		
DT	01-OCT-2000 (TREMREL. 15, Created)		
DT	01-OCT-2000 (TREMREL. 15, Last sequence update)		
DT	01-MAR-2003 (TREMREL. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Pestivirus.		
OX	NCBI_TaxID=11099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20323484; PubMed-10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	*Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.;		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RL	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL; AF268278; AAF82566.1; .		
DR	HSSP; P26863; IJXP.		
DR	MEROPS; S31.001; .		
DR	InterPro; IPR000280; CDvir_endptaseP80.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR004109; HCV NS3.		
DR	InterPro; IPR002166; HCV_RDRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR001005; Myb_DNA_Binding.		
DR	InterPro; IPR001568; RNase_T2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSVir.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF00271; Helicase_C; 1.		
DR	Pfam; PF00998; Viral_RDRP; 1.		

Q91RS1 hepatitis c  
Q91RQ8 hepatitis c  
Q91RT1 hepatitis c  
Q91R0 hepatitis c  
Q91R6 hepatitis c  
Q91R9 hepatitis c  
Q91RS3 hepatitis c  
Q91RS7 hepatitis c  
Q91R4 hepatitis c  
Q91RS8 hepatitis c  
Q91RT3 hepatitis c  
Q91RS5 hepatitis c  
Q91RS7 hepatitis c  
Q91RT0 hepatitis c  
Q91RS2 hepatitis c  
Q91RS6 hepatitis c  
Q91P61 hepatitis c  
Q91RS4 hepatitis c  
Q68533 hepatitis c  
Q91R7 hepatitis c  
Q91R6 hepatitis c  
Q36610 hepatitis c  
Q70817 hepatitis c  
Q91R8 hepatitis c  
Q70818 hepatitis c  
Q91R9 hepatitis c  
Q91R2 hepatitis c  
Q99AU2 hepatitis c

```

DR PRINTS: PR00729; COVENOPTASE.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICE; 1.
DR PROSITE: PS00037; MYB_1; 1.
DR PROSITE: PS00507; RDRP_POSITIVE; 1.
DR PROSITE: PS00521; RDRP_VIRAL; 1.
DR PROSITE: PS00531; RNASE_T2_2; 1.
DR PROSITE: PS00531; RNASE_T2_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase.
SQ SEQUENCE 4040 AA; 453073 MW; AD87791D05589DC CRC64;

Query Match 85.8%; Score 872.5; DB 12; Length 4040;
Best Local Similarity 89.28; Pred. No. 1e-75;
Matches 174; Conservative 5; Mismatches 13; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSDG---TAYAAQTGRGEGCOETSGTRDKKNQVEGEVQIVSTATQTFLAT 61
DB 10 GSVVIVGRIVLSSGSSITACAQQTGRLGCKITSLTGRDKKNQVEGEVQIVSTATQTFLAT 69
QY 62 SINGVLTVYHGAGTRTIAIPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCTCGSSDLVYL 121
DB 70 CINGVCTVYHGAGTRTIAIPKGPVTOMYTNVDQDLVGWPAQGSRLTPTCTCGSSDLVYL 129
QY 122 VTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAYGIFRAAVSTRGVAKA 181
DB 130 VTRHANYIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAYGIFRAAVSTRGVAKA 189
QY 182 VDFIPVESLETTRS 196
DB 190 VDFIPVENLETTRS 204

RESULT 2
Q36579 ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN STRAIN-H77;
RC SEQUENCE FROM N.A.
RX MEDLINE-97373636; PubMed-9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT *Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.*;
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR000745; HCV_NS4b.
DR InterPro; IPR001490; HCV_NS5a.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR007094; RNA_PSVir.

DR PF01543; HCV_capsid; 1.
DR PF01542; HCV_core; 1.
DR PF01539; HCV_env; 1.
DR PF01560; HCV_NS1; 1.
DR PF01538; HCV_NS2; 1.
DR PF02907; HCV_NS3; 1.
DR PF01006; HCV_NS4a; 1.
DR PF01001; HCV_NS4b; 1.
DR PF01506; HCV_NS5a; 1.
DR PF00271; helicase_C; 1.
DR PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW ATP-binding; coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0E809C63C1B9 CRC64;

Query Match 84.4%; Score 858.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 1.6e-74;
Matches 169; Conservative 10; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTGRGEGCOETSGTRDKKNQVEGEVQIVST 53
DB 1005 RRGQILLGPADGVMVSKGWRLAPITAYAAQTGRLGCKITSLTGRDKKNQVEGEVQIVST 1064
QY 54 ATQTFLATSLNGVLTVYHGAGTRTIAIPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCT 113
DB 1065 ATQTFLATSLNGVCTVYHGAGTRTIAIPKGPVTOMYTNVDQDLVGWPAQGSRLTPTCT 1124
QY 114 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAYGIFRAAV 173
DB 1125 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLYKSGSGGPLLCPAGHAYGIFRAAV 1184
QY 174 STRGVAKAVDFIPVENLETTRS 197
DB 1185 CTRGVAKAVDFIPVENLETTRS 1208

RESULT 3
Q81756 ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID-11103;
RN [1]
CC SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT *The nucleotide sequence of the Hepatitis C viral genome.*;
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27958; 1AIV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR007094; RNA_PSVir.
DR PF01560; HCV_NS1; 1.
DR PF01538; HCV_NS2; 1.

```

```

DR Pfam: PF02907: HCV_NS3; 1.
DR Pfam: PF01006: HCV_NS4a; 1.
DR Pfam: PF01001: HCV_NS4b; 1.
DR Pfam: PF01306: HCV_NS5a; 1.
DR Pfam: PF00271: helicase.C; 1.
DR Pfam: PF00998: Viral_RDRP; 1.
DR ProDom: PD186062: HCV_NS1; 1.
DR SMART: SM00487: DEXDC; 1.
DR PROSITE: PS05057: RDRP_POSITIVE; 1.
DR PROSITE: PS05052: RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT 2436
SQ SEQUENCE 2436 AA: 264734 MW: D7B9872900BE3125 CRC64;

Query Match 84.0%; Score 854.5; DB 12; Length 2436;
Best Local Similarity 82.8%; Pred. No. 2.9e-74;
Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVWIVGRIN-----LSGDTAYAAQOTRGECCOETSGTGRDKNQVEGEVQIVST 53
Db 555 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 614
QY 54 ATOTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCT 113
Db 615 AAQTFLATCINGVCWTYHGAGTRTIASPKGPVQMYTNVDKDLVGWQAPQGSRLTPTCT 674
QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 173
Db 675 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 734
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 735 CTRGVAKAVDFIPVENLETTMRSP 758

RESULT 4
Q9IFES ID Q9IFES PRELIMINARY; PRT; 3011 AA.
AC Q9IFES;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence."
RL J. Gen. Virol. 82:1291-1297(2001).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF271632; AAF8159.1; -.
DR HSSP: P27958; 1A1V.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.

InterPro: IPR002868; HCV_NS5a.
InterPro: IPR002166; HCV_RDRP.
InterPro: IPR001650; Helicase_C.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR007094; RNA_pol_PSVir.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NS1; 1.
Pfam: PF01538; HCV_NS2; 1.
Pfam: PF02907; HCV_NS3; 1.
Pfam: PF01006; HCV_NS4a; 1.
Pfam: PF01001; HCV_NS4b; 1.
Pfam: PF01506; HCV_NS5a; 1.
Pfam: PF00271; helicase.C; 1.
Pfam: PF00998; Viral_RDRP; 1.
ProDom: PD186062; HCV_NS1; 1.
SMART: SM00487; DEXDC; 1.
PROSITE: PS00190; CYTOCHROME_C; 1.
PROSITE: PS05057; RDRP_POSITIVE; 1.
PROSITE: PS05052; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA: 327124 MW: 2489CE74AC864E58 CRC64;

Query Match 84.0%; Score 854.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 3.9e-74;
Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVWIVGRIN-----LSGDTAYAAQOTRGECCOETSGTGRDKNQVEGEVQIVST 53
Db 1005 RRGREILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVST 1064
QY 54 ATOTFLATSIINGVLTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCT 113
Db 1065 AAQTFLATCINGVCWTYHGAGTRTIASPKGPVQMYTNVDKDLVGWQAPQGSRLTPTCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRPISYLYKSGSGGLPCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 5
Q9ELS8 ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus."
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF290978; AAG02099.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.

```





DR pfam: PF00998; Viral\_RDRP; 1.  
DR ProDom; PD186062; HCV\_NSI; 1.  
DR SMART; SM0487; DEXdc; 1.  
DR PROSITE; PS50507; RDRP\_POSITIVE; 1.  
DR PROSITE; PS50521; RDRP\_VIRAL; 1.  
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydrolase; Nonstructural protein; Polyprotein;  
KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ KQ SEQUENCE 3011 AA: 327112 MW: 087556B81CB5C198 CRC64:

Query Match	83.7%	Score	851.5	DB	12	Length	3011
Best Local Similarity	82.4%	Pred. No.	7.7e-74				
Matches	168	Conservative	10	Mismatches	17	Indels	9
Gaps	1						

  

QY	3	KKGSVVIVGRIN-----LGGDTAYAAQOTRGEGGCOET	SQTGRKKNQVEGEVQIVST	53
DB	1005	RGQEILGPGADGMVSKGWRLLAPITAYAAQOTRGLGCIIT	SLTGRDNQVEGEVQIVST	1064
QY	54	ATQTFLATSYINGVLWTVYHGGAGTGTIASPKGPVQWYTN	VDKDLVGMQAPGGSRLTPTCT	113
DB	1065	ATQTFLATCINGVCWTVYHGGAGTGTIASPKGPVQWYTN	VDQDLVGMWPAPOGSRSLTPTCT	1124
QY	114	CSSDLIYLVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSG	GPLLCPAGHAGVGFRAAV	173
DB	1125	CSSDLIYLVTRHADVIPVRRGDSRGSLLSPRISYLVKSGSG	GPLLCPAGHAGVGLFRAN	1184
QY	174	STRGVAKAVDFIPVESLETTMRSP	197	
DB	1185	CTRGVAKAVDFIPVENLGTMMRSP	1208	

  

RESULT	8
Q9PWX5	
ID	Q9PWX5
AC	PRELININARIY; PRT; 3015 AA.
DT	01-MAY-2000 (TReMBLrel. 13, Created)
DT	01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT	01-MAR-2003 (TReMBLrel. 23, last annotation update)
DE	Genome polyprotein.
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007099; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase.C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding: Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polypeptide;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BCIF190663A CRC64;

Query Match 83.7%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 7.7e-74;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 53
DB 1009 RRGQEILLGPADGMVSKGNWRLAPITAYAAQTRGLLGLTITSLTGRDKNOVEGEVQIVST 1068

QY 54 ATQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 113
DB 1069 ATQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 1128

QY 114 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLKSGSGGPGLLCPAGHAVGIFRAAV 173
DB 1129 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLKSGSGGPGLLCPAGHAVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTHMRSP 1212

RESULT 9
Q9PMW9 PRELIMINARY; PRT; 3015 AA.
AC Q9PMW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND M RNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase.C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase.C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding: Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polypeptide;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 83.7%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 7.7e-74;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDTAYAAQTRGEGCOETSGTRDKNOVEGEVQIVST 53
DB 1009 RRGQEILLGPADGMVSKGNWRLAPITAYAAQTRGLLGLTITSLTGRDKNOVEGEVQIVST 1068

QY 54 ATQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 113
DB 1069 ATQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPCT 1128

QY 114 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLKSGSGGPGLLCPAGHAVGIFRAAV 173
DB 1129 CGSSDLVLTVRHADVIPVRRGDSRGLSPRPISYLKSGSGGPGLLCPAGHAVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTTHMRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

```

```

RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.1Y;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369235; AAK54563.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 83.5%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 3.le-75;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Qy 19 TAYAAQOTRGEEGCOETSOTGRDNQVGEVQIVSTATQTFATSIINGVLTWYVHGAGTRT 78
Db 4 TAYAAQOTRGLLGCIIITSLTGRDNQVGEVQIVSTAQTFLATCINGVLTWYVHGAGTRT 63
Qy 79 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138
Db 64 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123
Qy 139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
Db 124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

RESULT 11
Q91RT5 ID Q91RT5 PRELIMINARY; PRT; 181 AA.
AC Q91RT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt. 4;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369218; AAK54543.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 83.5%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 3.le-75;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

Qy 19 TAYAAQOTRGEEGCOETSOTGRDNQVGEVQIVSTATQTFATSIINGVLTWYVHGAGTRT 78
Db 4 TAYAAQOTRGLLGCIIITSLTGRDNQVGEVQIVSTAQTFLATCINGVLTWYVHGAGTRT 63
Qy 79 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138
Db 64 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123
Qy 139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
Db 124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

```

```

Db 124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETTMRS 181

RESULT 12
Q91RR5 ID Q91RR5 PRELIMINARY; PRT; 181 AA.
AC Q91RR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.30;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369238; AAK54563.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match 83.3%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 4.9e-75;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 19 TAYAAQOTRGEEGCOETSOTGRDNQVGEVQIVSTATQTFATSIINGVLTWYVHGAGTRT 78
Db 4 TAYAAQOTRGLLGCIIITSLTGRDNQVGEVQIVSTAQTFLATCINGVLTWYVHGAGTRT 63
Qy 79 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 138
Db 64 IASPKGPVTQMTYNDKDLVGMQAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRGDSR 123
Qy 139 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
Db 124 GSLLSPRPISYLGSSGGPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 13
Q91RR2 ID Q91RR2 PRELIMINARY; PRT; 181 AA.
AC Q91RR2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369241; AAK54566.1; -.
DR InterPro; IPR004109; HCV_NS3.
DR Pfam; PF02907; HCV_NS3; 1.
KW Protease.
FT NON_TER 1
FT NON_TER 181
SQ SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-18  
Perfect score: 1017  
Sequence: 1 MKKKGWVIGRINLSGDTA.....YAKAVDFIPVSELTMTMSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=Cgn2\_1/USPTO.spool/US09965594/runat\_29082003\_151919\_28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN\_1\_1\_14686 -runat\_29082003\_151919\_28310 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pin.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	892.5	87.8	12734	6	ARI179057 Sequence
2	881.5	86.7	1998	6	ARI145264 Sequence
3	878.5	86.4	1998	6	ARI145268 Sequence
4	877.5	86.3	1998	6	ARI145262 Sequence
5	877.5	86.3	1998	6	ARI145263 Sequence
6	874.5	86.0	651	6	ARI145234 Sequence
7	874.5	86.0	1998	6	ARI145266 Sequence
8	874.5	86.0	1998	6	ARI145267 Sequence
9	873.5	85.9	1998	6	ARI145261 Sequence
10	873.5	85.9	2016	6	ARI145269 Sequence
11	872.5	85.8	12734	14	AF268278 Pestivirus
12	871.5	85.7	651	6	ARI145258 Sequence
13	870.5	85.6	651	6	ARI145252 Sequence
14	870.5	85.6	651	6	ARI145253 Sequence
15	870.5	85.6	1998	6	ARI145265 Sequence
16	870.5	85.6	2016	6	ARI145270 Sequence
17	870	85.5	648	6	ARI145274 Sequence
18	868	85.3	648	6	ARI145272 Sequence
19	867.5	85.3	651	6	ARI145256 Sequence
20	867.5	85.3	651	6	ARI145257 Sequence
21	867.5	85.3	651	6	ARI145260 Sequence
22	866.5	85.2	651	6	ARI145251 Sequence
23	866	85.2	648	6	ARI145273 Sequence
24	864	85.0	648	6	ARI145271 Sequence
25	863.5	84.9	651	6	ARI145255 Sequence
26	863.5	84.9	651	6	ARI145259 Sequence
27	861	84.7	8157	6	ARI127810 Sequence
28	861	84.7	8157	6	BD081911 Hepatitis
29	859	84.5	1932	6	ARI127809 Sequence
30	859	84.5	1932	6	BD081910 Hepatitis
31	858.5	84.4	9646	6	ARI110828 Sequence
32	858.5	84.4	9646	6	BD069982 Functiona
33	858.5	84.4	9646	14	AF009806 Hepatitis
34	858.5	84.4	12980	6	ARI110831 Sequence
35	858.5	84.4	12980	6	BD069985 Functiona
36	854.5	84.0	5360	6	ARI118686 Sequence
37	854.5	84.0	5360	6	I06434 Sequence 48
38	854.5	84.0	5360	6	I09328 Sequence 8
39	854.5	84.0	5360	6	ARI118692 Sequence
40	854.5	84.0	6785	6	I06440 Sequence 54
41	854.5	84.0	6785	6	I09329 Sequence 10
42	854.5	84.0	7310	6	ARI118696 Sequence
43	854.5	84.0	7310	6	I09331 Sequence 15
44	854.5	84.0	7310	14	HPCPOLYP M32084 Hepatitis C
45	854.5	84.0	8316	6	ARI118703 Sequence

ALIGNMENTS

RESULT 1

ARI179057  
LOCUS ARI179057 12734 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6326137.  
ACCESSION ARI179057  
VERSION ARI179057.1 GI:20220612  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 12734)  
AUTHORS Hong, Z., Lai, V.C.H. and Lau, J.Y.N.  
TITLE Hepatitis C virus protease-dependent chimeric pestivirus  
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;  
FEATURES Location/Qualifiers  
source  
1..12734  
/organism="unknown"  
BASE COUNT 4032 a 2604 c 3295 g 2803 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1..1e-65 Length: 12734  
Score: 892.50 Matches: 177  
Percent Similarity: 92.8% Conservative: 4  
Best Local Similarity: 90.7% Mismatches: 11  
Query Match: 87.7% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x ARI179057 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
DB 413 GGTAGTGGTGTATTTGGTGGTAGAATATTTTCTGTGTAGTGTAGTATCATCGGCTAC 472  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
DB 473 GCCAGCAGCAGAGGCTCTCTAGGGGTAAAGATCACCACTGCTGACTGGCGGGACAAA 532  
QY 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 533 AACCAAGTGGAGGGTGGAGTCCAGATGCTGTCAACTGCTACCAACCTTCTCTGGCAACG 592  
QY 62 SerIleAsnGlyValIleuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 593 TGCATCAATGGGCTATGCTGGACTCTCTACCAAGGGCCGGAACAGGACCATCGCATCA 652  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101  
DB 653 CCCAAGGGTCTCTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGCTGGCCC 712  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 713 GCTCTCAAGGTTCGGCTCATGTACACCTTGACCTCGGCTCTCTGGACCTTTACTTG 772  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141  
DB 773 GTTACAGGACGCGCAGCTCATTTCCCTGGCGCGGAGGTGATAGCAGGGTAGCCTG 832  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
DB 833 CTTTCGCCCGGCCCATTTCTACCTAACAGGCTCTCGGGGGTCCGCTGTGTGGCCCC 892  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 893 GCGGGACAGCGCTGGGCTATTACAGGCGCGGTGTGCACCCGCTGGAGTGGCCAAAGCG 952  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGACAAACCATGAGATCC 997

RESULT 2  
ARI145264  
LOCUS ARI145264 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 105 from patent US 6211338.

ACCESSION ARI45264  
VERSION ARI45264.1 GI:15107131  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1998)  
AUTHORS Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 105 03-APR-2001;  
FEATURES Location/Qualifiers  
source  
1..1998  
/organism="unknown"  
BASE COUNT 411 a 595 c 569 g 423 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1..1e-65 Length: 1998  
Score: 881.50 Matches: 167  
Percent Similarity: 93.37% Conservative: 16  
Best Local Similarity: 85.20% Mismatches: 10  
Query Match: 86.68% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x ARI45264 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
DB 64 GGTCTGTGTGTATTTGGTGGTAGAATATTTTCTGTGTAGTGTAGTATCATCGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
DB 124 TCCCAACAGACGCGGGCTACTTGGTTGCAAGAAGACTAGCCTTACAGCGCGGACAAAG 183  
QY 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 184 AACCAAGTGGAGGGTGGAGTCCAGTGTGTTCACCGCAACAACTCTCTCTGGCAAC 243  
QY 62 SerIleAsnGlyValIleuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 244 TCGCTCAACGCGGTGTGTGGACCTTTACCATGTGTGTGGCTCAAGACCTTAGCCGCG 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101  
DB 304 CCMAAGGGGCCATGCCAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 364 GCGCCCGCGGGCGGTCTCTTGTACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141  
DB 424 GTCAGAGACATGCTGACGTCTATTCGGTGGCGGGCGGCGGACAGTAGGGGAGCCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
DB 484 CTTCCCGCCAGCGCTCTCTCTTGAAGGCTCTTGGGCTGTGTCCACTGCTGCTGCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 544 TCGGGGACGCTGTGGCATCTTCGGCTTCCGCTATGACCCGGGGGGTGTGCAAGGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAATACTATGCGGTCTCCG 651

RESULT 3  
ARI45268  
LOCUS ARI45268 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 109 from patent US 6211338.  
ACCESSION ARI45268  
VERSION ARI45268.1 GI:15107135

## KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

Patent: US 6211338-A 109 03-APR-2001;

FEATURES Location/Qualifiers

1..1998

/organism="unknown"

BASE COUNT 411 a 595 c 569 g 423 t

ORIGIN

Alignment Scores:

Pred. No.: 1.99e-65 Length: 1998  
 Score: 878.50 Matches: 166  
 Percent Similarity: 93.37% Conservative: 17  
 Best Local Similarity: 84.69% Mismatches: 10  
 Query Match: 86.38% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145268 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTCTGTTGTTATTGTTGGTAGAATTATTTATCTGGTAGTGTAGTATCATCGGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 :|||:|||||  
 Db 124 TCCCAACAGACGCGGGCCCTACTTGGTTGCAAGAACTAGCCTTACAGGCGGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 :|||:|||||  
 Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGGTTTCCACCCACACAACTCTCTCGGGGACC 243  
 QY 62 SerIleAsnGlyValLeuTyrPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 :|||:|||||  
 Db 244 TCGCTCAACGGCGTGTGTGGACCGTTTACCATGTGCTGGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 :|||:|||||  
 Db 304 CCNAAGGGCCCAATCACCAGCATGTACACTAATGTGCACGACCTCTCGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 :|||:|||||  
 Db 364 CGCGCCCGCGGGCGGCTCTTCCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 :|||:|||||  
 Db 424 GTCAGAGACATGCTGCTCTACTTGAAGGGCTCTGCTGGTGTCCACTGCTCGCCCT 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 :|||:|||||  
 Db 484 CTCCTCCCGCCAGCGCTGCTCTCTACTTGAAGGGCTCTTCCGGTGTCCACTGCTCGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 :|||:|||||  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTGGCAAGGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 :|||:|||||  
 Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACACTACTATGCGGCTCCG 651

## RESULT 4

AR145262

LOCUS

DEFINITION

Sequence 103 from patent US 6211338.

ACCESSION

AR145262

VERSION

AR145262.1

KEYWORDS

Source

Unknown.

AR145262 1998 bp DNA linear PAT 08-AUG-2001

Sequence 103 from patent US 6211338.

ACCESSION AR145262

VERSION AR145262.1

GI:15107129

Unknown.

## ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

Patent: US 6211338-A 103 03-APR-2001;

FEATURES Location/Qualifiers

1..1998

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 2.42e-65 Length: 1998  
 Score: 877.50 Matches: 167  
 Percent Similarity: 92.86% Conservative: 15  
 Best Local Similarity: 85.20% Mismatches: 11  
 Query Match: 86.28% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145262 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTCTGTTGTTATTGTTGGTAGAATTATTTATCTGGTAGTGTAGTATCATCGGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 :|||:|||||  
 Db 124 TCCCAACAGACGCGGGCCCTACTTGGTTGCAAGAACTAGCCTTACAGGCGGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 :|||:|||||  
 Db 184 AACCAAGTCGAGGAGAGGTTTCAGTGGTTTCCACCCGACACAACTCTCTCGGGGACC 243  
 QY 62 SerIleAsnGlyValLeuTyrPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 :|||:|||||  
 Db 244 TCGCTCAACGGCGTGTGTGGACCGTTTACCATGTGCTGGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 :|||:|||||  
 Db 304 CCNAAGGGCCCAATCACCAGCATGTACACTAATGTGCACGACCTCTCGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 :|||:|||||  
 Db 364 CGCGCCCGCGGGCGGCTCTTCCACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 :|||:|||||  
 Db 424 GTCAGAGACATGCTGCTGCTACTTGAAGGGCTCTTCCGGTGTCCACTGCTCGCCCT 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 :|||:|||||  
 Db 484 CTCCTCCCGCCAGCGCTGCTCTCTACTTGAAGGGCTCTTCCGGTGTCCACTGCTCGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 :|||:|||||  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTGGCAAGGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 :|||:|||||  
 Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACACTACTATGCGGCTCCG 651

## RESULT 5

AR145263

LOCUS

DEFINITION

Sequence 104 from patent US 6211338.

ACCESSION

AR145263

VERSION

AR145263.1

KEYWORDS

Source

Unknown.

ORGANISM

Unclassified.

AR145263 1998 bp DNA linear PAT 08-AUG-2001

Sequence 104 from patent US 6211338.

ACCESSION AR145263

VERSION AR145263.1

GI:15107130

Unknown.

Unclassified.





source	1. .1998	/organism="unknown"
LOCATIONS		
LOCATIONS/Qualifiers		

BASE COUNT 409 a 597 c 567 g 425 t  
ORIGIN

## Alignment Scores:

Pred. No.: 5,29e-65 Length: 1998  
Score: 873.50 Matches: 167  
Percent Similarity: 92.35% Conservative: 14  
Best Local Similarity: 85.20% Mismatches: 12  
Query Match: 85.89% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21  
DB 64 GGTCTGTGTATGTTGGTAGAATATTATTCCTGAGTGTAGTATCATCGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
DB 124 TCCCAACAGACGCGGGCCCTACTTGGTTCATCATCATCTAGCCTTACAGCGCGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 184 AACCAAGTCGAGGAGAGGTTTCAGGTGGTTCACCGCAACACAAATCCTTCCCTGGCGGACC 243  
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 244 TCGGTCAACGCGGTGTGTGACCGTTTACCNTGGTGGCTCAAGACCTTAGCGCGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 304 CCAAGGGGCGCAATCACCCAGATACACTAATGTGGACCAAGACCTTCGCGGTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 364 GCGCCCGCGGGCGCGTCTTTCACACCATGACCTGTGGCGGCGGCGACAGTAGGGGAGCGTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
DB 424 GTACAGACATGCTGACGTCATTCCTGAGGCGCTTGAAGGCGCTTCGGGTGCTCCACTGCTCGCCCT 543  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161  
DB 484 CTCTCCCGCAGCGCTCTCTCTACTTGAAGGCGCTTCGGGTGCTCCACTGCTCGCCCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 544 TCGGGGACGCTGTGGCATCTTCGGGCTCGCGTATGCACCGCGGGGTTGCCAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGlnThrThrMetArgSerPro 197  
DB 604 GTGGACTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 10  
LOCUS AR145269 2016 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 110 from patent US 6211338.  
ACCESSION AR145269  
VERSION AR145269.1 GI:15107136  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE  
1 (bases 1 to 2016)  
AUTHORS Malcolim.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 110 03-APR-2001;  
FEATURES Location/Qualifiers  
source /organism="unknown"  
1. .2016

BASE COUNT 412 a 603 c 570 g 431 t  
ORIGIN

## Alignment Scores:

Pred. No.: 5,35e-65 Length: 2016  
Score: 873.50 Matches: 167  
Percent Similarity: 92.35% Conservative: 14  
Best Local Similarity: 85.20% Mismatches: 12  
Query Match: 85.89% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21  
DB 82 GGTCTGTGTATGTTGGTAGAATATTATTCCTGAGTGTAGTATCATCGGCTAC 141  
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41  
DB 142 TCCCAACAGACGCGGGCCCTACTTGGTTCATCATCATCTAGCCTTACAGCGCGGACAAG 201  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 202 AACCAAGTCGAGGAGAGGTTTCAGGTGGTTCACCGCAACACAAATCCTTCCCTGGCGGACC 261  
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 262 TCGGTCAACGCGGTGTGTGACCGTTTACCNTGGTGGCTCAAGACCTTAGCGCGC 321  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 322 CCAAGGGGCGCAATCACCCAGATGACACTAATGTGGACCAAGACCTTCGCGGTGGCAG 381  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 382 GCGCCCGCGGGCGCGTCTTTCACACCATGACCTGTGGCGGCTCAGACCTTACTTG 441  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
DB 442 GTACAGACATGCTGACGTCATTCCTGAGGCGCTTGAAGGCGCTTCGGGTGCTCCACTGCTCGCCCT 501  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyGlyProLeuLeuCysPro 161  
DB 502 CTCTCCCGCAGCGCTCTCTCTACTTGAAGGCGCTTCGGGTGCTCCACTGCTCGCCCT 561  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 562 TCGGGGACGCTGTGGCATCTTCGGGCTCGCGTATGCACCGCGGGGTTGCCAAGCG 621  
QY 182 ValAspPheIleProValGluSerLeuGlnThrThrMetArgSerPro 197  
DB 622 GTGGACTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 669

RESULT 11  
LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000  
DEFINITION Pestivirus type 1, complete genome.  
ACCESSION AF268278  
VERSION AF268278.1 GI:9049956  
KEYWORDS  
SOURCE Pestivirus type 1  
ORGANISM Pestivirus type 1  
Viruses: ssRNA positive-strand viruses, no DNA stage: Flaviviridae;  
Pestivirus.

REFERENCE  
1 (bases 1 to 12734)  
AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassiliev,V.,  
Donis,R.O., Hong,Z. and Lau,J.Y.  
TITLE Generation and characterization of a hepatitis C virus NS3  
protease-dependent bovine viral diarrhoea virus  
JOURNAL J. Virol. 74 (14), 6339-6347 (2000)  
MEDLINE 20323484  
PUBMED 10864644  
REFERENCE  
2 (bases 1 to 12734)  
AUTHORS Lai,V.C.H. and Hong,Z.  
TITLE Direct Submission

JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

# FEATURES

```

source
1. .12734
   /organism="Pestivirus type 1"
   /mol_type="genomic RNA"
   /db_xref="taxon:11099"
1. .385
386. .12508
   /codon_start=1
   /product="polyprotein"
   /protein_id="AAP82566.1"
   /db_xref="GI:9049957"
   /translation="MELNTEGSGSVIVGRVLGSGSITACAOOTCLLGCKITSL
TGRKQVEGEVQIVSTQATOTLATCINGCYVTVHGAGTRIAPKGPVQIYNTVD
QDLVGPAPGSGSLTCTCGSDSLVTRHANVIVRRGDSRLSSPRISLYLKG
SSGGLPCFAGHAGVFRACVTRGVAKAVDFIPVENETTRSGGADVEDVCCSM
SYDTEKATKKTQPORLERGKMKIVPKESKDKTKPPDATIVVEGVYQVRKK
GRTKNTQDGLVHNKPKQSKLEKALAWAIIAIVLFOVTMGENTQWNLQDNG
TEGIORAMFORGVNRSLSHGIMPEKICTGVPSHLATDIELKTIGHMDASEKNTYCCR
LORHNKKGWCKNYNIEPKILLMNTOANLTGEGOPRECAVTCYDRASDLNVVTOA
RDSPTLTCGCKKNSFAGILLMRGCFNFEIAASDVLFKEHERISMFQDTLLYDGL
TNSLEGARQGTAKLTWLKGQGLIGLKLENSKTFWFGAIAASPYCDVDKIGIYWT
KNCPTACLPKNTKIVGKPFDAEDGKILHEMGHLSVLLSLVLSLDFAPETASV
MYLILHFSIPQSHVDVMDKLTQNLNLTVELTADVIPGSVNLGKWICRPNMWPYET
TVLVAEEVSQVYKVLRAIRLRLTRWNAATTAFILCLVIVRGVQVQGLWLLIT
GVGHLDCKPERSVYAJAKDERIGOLGAELTITWKEYSPGKLEDTMTWAMEDCKLM
YLQCRTRRYLALHTRALPVSVEFKLFDGRKQEDVEMNDNEFGLPCDAPKIV
RGENTILLNGPAFQVCPVIGWTVGTSCTSFNMDLTLATVTVTYRSPKPFPHRQCII
QKNLGEDHNCILGNNWTCVPGDQLLYKGSIESCKWQGTQPKSEBGLPHYPICKCL
ENETGVACTFNTYTKTKNKPYPROSFYQOYLKGEYQYWFDEVTDRHDPYAEIS
EGVEXKATFNTYTKTKNKPYPROSFYQOYLKGEYQYWFDEVTDRHDPYAEIS
LVVVALESQVYKVLRAIRLRLTRWNAATTAFILCLVIVRGVQVQGLWLLIT
LYLLREESKVVKLLYHLYVHPKTSVIVLIMIGDVVKADSGQEVYLGKIDICFT
TVVILVIGLILARDEPTVPLVIMAAIBVTELTQPGVDYIAVAVMTLILMVSYTD
YFRKKNLOCLISLVSQVFLRLSYLIGRIEMPEVTPINWRPLTLLILLLISTIVTR
WKVDVALQCCPILLVITLMADEFTLILITVELYKLYLTKVTRDIERSLWGG
IDTVRVDYVDVDESEGVYLPFSRQAKQNFSLPLIKATLISCVSKWQIYMSY
LITDFMYHRRKYIELISGNTIISRLVANIELNMSMEEEKLKFYLLSGLRLN
LIHKVRNVTASVTEGEVYKPKIMTIKASTLSKRNHICITVCBGRWKGTGTC
PKGRIKPTICNSLADFEERYHKRIFIREGNEFGMCSRCOGKRRRPFMDREKASR
YCAEENRHPAESGDFWAESSLGLKITYPALMDGVYDITWACQORVIGSPDHRV
PCHISGRSNPFQETNGFYVARGOLFLPVLATKVMMLVNGGEEIGNLEHL
GWLIRGPAVCKTIEHEKHNILDKLTAFPGIMPTTPRAPRYPVSLLRKGLLE
TGWAYTHQGISVHVATAGKDLVCSMGRTRVQCNNRLTDEYGVKTDSCPD
GARYLNPANVYISGKAGVHLQGTGTECTVAGTAPDFDLKNGWSGLPIFE
ASSGRVYGVKGNKESKPTKIMSGITQVSKNTADLTWKKITSMNRGDKQITLA
TGAKTTELPAKVEIEIGRHKRVLLIPLRAAAESVYQVRLKHPISPLNRLGDMKE
GDMATGYASYGVFCOMPQKPLAAWVEYSYIFLDEYHCATPEOLAIIGKIHRES
IRVAMTATPAGSVTTIGQKHPIEFIEPVMKGEDLSQFLDIAGLKIPVDMKGM
LVFVPTNMAVEVAKKLLKAGVNSGIYISGEDPANRVVISQSPYVAINAIESGV
LPDLDTVIDTLCKEKRVRVSSKIPFIVTGLKMAVTVGEQARRGRVGRVPRYIR
SOETATGSKDYHDLQARIGLEDINTVTSFRENHNDMSLYEEDSLITOLEINN
LLISEDLPAAVKIMARTDHPEDIQIAYNSVEQVPLPKRNGEVTVTYENYSFLN
ARKLGEDVYVIVATDEEDLVLGLDWPDPGNQOVVETKALQVOTGLSNAENALL
VALGVYGOALSRAVPMVTDIYTEDORLEDTTHLOYAPNAIKDTGETELKELAS
GDVEKINGALSIDVAAAGLEVSKAEEKITAPLKENAANKGVYKQVFDLSLENKEE
IIRGLMGTHTALYSIAARLGHETAFATLKLWAFPGSVSDHVKQKAAVDLVYIV
HNKSPFGESEVTOQEGRRFVASLFSALATYTKTWNHLSKVEPALAYLPATAS
LKMPFTPLRESVYLTSTYIKYIKGKSDGLGTGISAAMEILSNQNPVSGVISMV
LGVCAIAAHNAIESSQKRTLKLVFNKPLDQAATDELVKENPEKIMALPEAQVTI
GNPLRLIYHLYGYVYKGMWEAKELSERAGNLTFLIMFAELGLMDSGOKIRNLSGN
YILLDYLGNKGLKMWLGRAPSCDPTSDERILPTONYLAVETRCQGY
EMAKRNVNGKILKVEESGFFLCRNPGRGPVNVYRKYIDNLRINLFAVLEQOVE
HYTKGVAKIDYSGKMLLADKWEVHGVTIRLAKRYTVGVGFNGAYLGDEPNHRAV
ERDATTIKNTVQFLAKMKGCAFTYDLTISNLTRELVRHNNLEKETPTATVTLW
AYTFVNDVGTIKPVLGRVYIPDPVVDINLQPEVQDVTSEVGTIIGRELTMTGVP
VLEKEVPADSDNONSVKIGLDEGNYPGPIQTHLTEEIHNRDARPPILIGSRNLS
NRKATIRNLYTGNDRPREDLMAAGRLVVALRDVDPLESEMDFKGTFLDRALE
ALSGLQPKQVTKAVARNLIEQKQVEIPNPFASDDPVELEVALKNKTYILVDGVE
VKDQAKALGATYLAQGNWPELGGVHLGTIPARRVKIHPEYAYLKLDFIEEEKKPR
KSNKGNHASAYLAQGNWPELGGVHLGTIPARRVKIHPEYAYLKLDFIEEEKKPR
VKDVTIREHNKWLKIRFOGNLNTKKMLNPGKLSBQDLREGKRNINHQITGTMSS

```

```

AGIRLEKLPVRAQCTDTKTTFHEAIRDKIDKSENQNPELNKLLLELPHITIAOPTLKH
YGEVTEQLEAGINRKGAAFFLEKNKIGEVDSERHLVSEQLVRLDKAGRKIYETAI
PKREKRVSDWQAGDLVVEKPRVIOYPEAKTRLAITKVMYNNWVQPVVPGYEGK
TPLFNIDKVRKEMWDFNEKPRVSDFTKAWDTQVTSKOLQIGELQKYAFKEMHFI
DTIDRMTVEVPTADGEVYIRNGQSGOPOTSAGNSMLNLTMMYAFCESTGVYK
SPNRVARIHVCDDGFLITEKGLGAFANKQIILHEACKPOKITTEGEMKVAYRFED
IECSHTPPVPRMSDNTSSHMAGRD/AVILSKWATRLDSSGSGRTATKAVAFSFL
MYSWNLVRRICILLVLSQFPETDPSKHAITYYKGDPIGAIKDVIGRNSLSEKARTGEK
LANLNLSTLGTWTKRITKRIIDCAVIGKEEKNLWADRLISGRSLHLYIPDKGF
TLOGKHVEQLQRLTETNPVMGVGTERYKLGPIVNLRLRLKILLMTAVGVS"
3'UTR
BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others
ORIGIN
Alignment Scores: 5.58e-64 Length: 12734
Pred. No.: 872.50 Matches: 174
Score: 872.50 Matches: 174
Percent Similarity: 91.79% Conservative: 5
Best Local Similarity: 89.23% Mismatches: 13
Query Match: 85.79% Indels: 3
DB: 14 Gaps: 1
US-09-965-594-18 (1-197) x AF268278 (1-12734)
QY 5 GlySerValValIleValGlyAtrGileAsnLeuSerGlyAsp-----ThrAlaTyr 21
413 GGTAGTGTGTTATTGTTGGTAGAATTTTTCCTGGTAGTGTAGTATCACGGCGTGC 472
Db
QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41
473 GCCCAGCAGCAGCAGAGAGCCCTCTTAGGGTGTAAAGTACCATCTGACTGGCGCGGCAAA 532
Db
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
533 AACCAAGTGAGGGTGAGGTGAGGTCCAGATCGTCACTGCTACCCAAACCTTCTCGGCAAG 592
Db
QY 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81
593 TGCATCAATGGGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652
Db
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
653 CCCAAGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712
Db
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
713 GCTCCTCAAGGTCTCCGCTCATTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772
Db
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
773 GTTACGAGGACACGCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832
Db
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
833 CTTTCGCGCGCGCCCATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 892
Db
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
893 GCGGACACGCGCGCGCTATTACGCGCGCGCGCTGTCACCCGCGCGCTGTCAGTGGCGGCGG 952
Db
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
953 GTGACTTTATCCCTGTGGAGAACCTAGACACCAACACGAGATCC 997
Db
RESULT 12
ARI45258
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

## Unclassified.

## REFERENCE

1 (bases 1 to 651)

## AUTHORS

Malcolm.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.

## TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

## JOURNAL

protease and NS4A cofactor peptide

## FEATURES

Patent: US 6211338-A 93 03-APR-2001;

## source

Location/Qualifiers

## BASE COUNT

1. 651

## ORIGIN

/organism="unknown"

120 a 187 c 200 g 144 t

## Alignment Scores:

Pred. No.: 2,12e-65 Length: 651

Score: 871.50 Matches: 165

Percent Similarity: 93.33% Conservative: 17

Best Local Similarity: 84.62% Mismatches: 10

Query Match: 85.69% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x ARI45258 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Db 64 GGTTCGTGTTGTTATGTTGTTAGATATTTTATCTGTTAGTGGTAGTATCACGGCTTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGCGGGGCTACTTGGTTCGACAGAGACTAGCTTACAGCGGGGACAAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACACAGTCGAGGAGAGGTTTCCAGTGGTTTCCACCGCAACACATCTCTCTGCGGACC 243

QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TCGCTCAACGGCGTGTGGACGTTTACCATGTGCTGGCTCAAGACCTTAGCCGGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101

Db 304 CCAAGGGGCCAATCCACAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 GCCTCCCGCGGGCGGCTTCCCTTGACACCATGCATCTGGCGAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATCTGACGCTCATTCGGTGCCTGCGCGGGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCTCCCGCGGGCGGCTGCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGCTCTTCGGGCTGCGGTATGCACCGGGGGGTTTCGGAAGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196

Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 13

ARI45252

LOCUS

Sequence 93 from patent US 6211338.

ACCESSION

ARI45252

VERSION

ARI45252.1 GI:15107119

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 651)

Malcolm.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.

Single-chain recombinant complexes of hepatitis C virus NS3

## AUTHORS

Malcolm.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.

## TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

## JOURNAL

protease and NS4A cofactor peptide

## FEATURES

Patent: US 6211338-A 93 03-APR-2001;

## source

Location/Qualifiers

## BASE COUNT

1. 651

## ORIGIN

/organism="unknown"

119 a 188 c 199 g 145 t

## Alignment Scores:

Pred. No.: 2,58e-65 Length: 651

Score: 870.50 Matches: 166

Percent Similarity: 92.82% Conservative: 15

Best Local Similarity: 85.13% Mismatches: 11

Query Match: 85.59% Indels: 3

DB: 6 Gaps: 1

US-09-965-594-18 (1-197) x ARI45252 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21

Db 64 GGTTCGTGTTGTTATGTTGTTAGATATTTTATCTGTTAGTGGTAGTATCACGGCTTAC 123

QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41

Db 124 TCCCAACAGCGGGGCTACTTGGTTCGACAGATCCTAGCTTACAGCGGGGACAAAG 183

QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61

Db 184 AACACAGTCGAGGAGAGGTTTCCAGTGGTTTCCACCGCAACACATCTCTCTGCGGACC 243

QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81

Db 244 TCGCTCAACGGCGTGTGGACGTTTACCATGTGCTGGCTCAAGACCTTAGCCGGC 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101

Db 304 CCAAGGGGCCAATCCACAGATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363

QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121

Db 364 GCCTCCCGCGGGCGGCTTCCCTTGACACCATGCATCTGTGGCAGCTCAGACCTTTACTTG 423

QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141

Db 424 GTCACGAGACATCTGACGCTCATTCGGTGCCTGCGCGGGGCGACAGTAGGGGAGCCTG 483

QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161

Db 484 CTCTCCCGCGGGCGGCTGCT 543

QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

Db 544 TCGGGGACGCTGTGGGCTCTTCGGGCTGCGGTATGCACCGGGGGGTTTCGGAAGCG 603

QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196

Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 14

ARI45253

LOCUS

Sequence 94 from patent US 6211338.

ACCESSION

ARI45253

VERSION

ARI45253.1 GI:15107120

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 651)

Malcolm.B.A., Taremi.S.Shane., Weber.P.C. and Yao.N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide  
Patent: US 6211338-A 94 03-APR-2001;

## JOURNAL

FEATURES

Location/Qualifiers

source

1. .651

BASE COUNT

119 a 188 c 199 g 145 t

ORIGIN

/organism="unknown"

Alignment Scores:

Pred. No.:

2.58e-65

Score:

870.50

Percent Similarity:

92.82%

Best Local Similarity:

85.13%

Query Match:

85.59%

DB:

6

Length:

651

Matches:

166

Conservative:

15

Mismatches:

11

Indels:

3

Gaps:

1

US-09-965-594-18 (1-197) x AR145253 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTAGTAGTATACACGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCACACAGACGGGGGCTACTTGGTTCATCAAGACTAGCCTTACAGGCGGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTCCACCGCAACAACTCTTCCTGGCGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGCTCAACGGCGTGTGTGGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACCAAGACCTCGTCGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGGCCCCGGCGCGCTTCCTTGACACCATGCACTGTGGACCTCAGACCTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGACACATGCTGACGTCATTCGGTGGCGGGCGGACAGTAGGGGAGCGCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGCAGCGCTGCTCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCCGTATGCCACCGGGGGTTCGAAGGCG 603  
 QY 182 ValAspPheIleProValIleSerLeuGluThrThrMetArgSer 196  
 DB 604 GTGGACTTGTGCGGTAGATCCATGGAACTACTATGCGGTCT 648

## RESULT 15

AR145265

LOCUS

DEFINITION

Sequence 106 from patent US 6211338.

ACCESSION

AR145265

VERSION

AR145265.1

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 1998)

AUTHORS

Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 106 03-APR-2001;

## FEATURES

Location/Qualifiers

1. .1998

/organism="unknown"

BASE COUNT

409 a 597 c 567 g 425 t

ORIGIN

Alignment Scores:

Pred. No.:

9.53e-65

Score:

870.50

Percent Similarity:

92.35%

Best Local Similarity:

84.69%

Query Match:

85.59%

DB:

6

Length:

1998

Matches:

166

Conservative:

15

Mismatches:

12

Indels:

3

Gaps:

1

US-09-965-594-18 (1-197) x AR145265 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTAGTAGTATACACGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCACACAGACGGGGGCTACTTGGTTCATCACTACCTTACAGCCTAGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTCCACCGCAACAACTCTTCCTGGCGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGCTCAACGGCGTGTGTGGACCGTTTACCATGGTCTGGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACCAAGACCTCGTCGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGGCCCCGGCGCGCTTCCTTGACACCATGCACTGTGGACCTCAGACCTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGACACATGCTGACGTCATTCGGTGGCGGGCGGACAGTAGGGGAGCGCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGCAGCGCTGCTCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCCGTATGCCACCGGGGGTTCGAAGGCG 603  
 QY 182 ValAspPheIleProValIleSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTGTGCGGTAGAGTCCATGGAACTACTATGCGGTCTCGG 651

Search completed: August 31, 2003, 00:46:23

Job time : 2569.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-18

Perfect score: 1017

Sequence: 1 MKKGSVVIVCRINLSGDTA.....VAKAVDFIPVESLETHMRSP 197

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlp  
-O=/cqn2.1/USPTO\_pool/US09965594/runat\_29082003\_151918\_28302/app\_query.fasta\_1.2872  
-DB=N\_Geneseq\_19Jun03 -QWTF=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -RCGN\_1\_1412.ernat\_29082003\_151918\_28302 -NCPU=6 -ICPU=3  
-NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1017	100.0	594	21	AAA73332	Hepatitis C virus
2	1005	98.8	594	21	AAA73333	Hepatitis C virus
3	1002	98.5	594	21	AAA73331	Hepatitis C virus
4	995	97.8	594	21	AAA73334	Hepatitis C virus
5	985	96.9	594	21	AAA73330	Hepatitis C virus
6	951	93.5	588	21	AAA73329	Hepatitis C virus
7	946	93.0	594	21	AAA73335	Hepatitis C virus
8	912	89.7	588	21	AAA73328	Hepatitis C virus
9	892.5	87.8	12734	24	ABA95615	Chimeric BVDV/HCV
10	881.5	86.7	1998	20	AAH80355	HCV NS4A-NS3 compl
11	878.5	86.4	1998	20	AAH80359	HCV NS4A-NS3 compl
12	877.5	86.3	1998	20	AAH80353	HCV NS4A-NS3 compl
13	877.5	86.3	1998	20	AAH80354	HCV NS4A-NS3 compl
14	874.5	86.0	612	25	ABX15706	Anti-viral synthet
15	874.5	86.0	651	20	AAH80345	HCV NS4A-NS3 compl
16	874.5	86.0	1998	20	AAH80357	HCV NS4A-NS3 compl
17	874.5	86.0	1998	20	AAH80358	HCV NS4A-NS3 compl
18	873.5	85.9	1998	20	AAH80352	HCV NS4A-NS3 compl
19	873.5	85.9	2013	20	AAH80360	HCV NS4A-NS3 compl
20	871.5	85.7	651	20	AAH80349	HCV NS4A-NS3 compl
21	870.5	85.6	651	20	AAH80343	HCV NS4A-NS3 compl
22	870.5	85.6	651	20	AAH80344	HCV NS4A-NS3 compl
23	870.5	85.6	1998	20	AAH80356	HCV NS4A-NS3 compl
24	870.5	85.6	2016	20	AAH80361	HCV NS4A-NS3 compl
25	870	85.5	648	20	AAH80365	HCV NS4A-NS3 compl
26	868	85.3	648	20	AAH80363	HCV NS4A-NS3 compl
27	867.5	85.3	650	20	AAH80347	HCV NS4A-NS3 compl
28	867.5	85.3	651	20	AAH80348	HCV NS4A-NS3 compl
29	867.5	85.3	651	20	AAH80351	HCV NS4A-NS3 compl
30	866.5	85.2	651	20	AAH80342	HCV NS4A-NS3 compl
31	864	84.9	648	20	AAH80362	HCV NS4A-NS3 compl
32	863.5	84.9	650	20	AAH80346	HCV NS4A-NS3 compl
33	863.5	84.9	651	20	AAH80350	HCV NS4A-NS3 compl
34	861	84.7	8145	20	AAH23259	Plasmid pET-BS(+)/
35	859	84.5	1933	20	AAH23258	HCV NS3 DNA. Hepa
36	858.5	84.4	9646	19	AAH59381	Hepatitis C virus
37	858.5	84.4	9646	24	ABK87285	cdNA encoding hepa
38	858.5	84.4	12980	19	AAH59364	Hepatitis C virus
39	858.5	84.4	12980	24	ABK87286	Hepatitis C virus
40	858.5	84.4	16622	21	AAZ36212	Nucleotide sequenc
41	854.5	84.0	5300	10	AAH92097	Combined open read
42	854.5	84.0	5360	10	AAH90327	Hepatitis C virus
43	854.5	84.0	6905	10	AAH92103	Combined open read
44	854.5	84.0	7310	10	AAH92106	Combined open read
45	854.5	84.0	7310	10	AAH90336	Composite hepatitis

ALIGNMENTS

RESULT 1  
AAA73332  
ID AAA73332 standard; DNA; 594 BP.  
XX  
AC AAA73332;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein; ds.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
FH Key Location/Qualifiers



```

FT CDS 1..594
PT /*tag= a
XX /product= "NS4A-NS3 fusion protein #5"
XX WO2000040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 99US-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
XX P-PSDB; AAB15223.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 15; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
XX Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1 34e-86 Length: 594
XX Score: 1017.00 Matches: 197
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-965-594-18 (1-197) x AAA73332 (1-594)
XX
XX 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
XX 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCGCGTATCAACCTGTCGGGTGACACCGCT 60
XX
XX 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
XX 61 TAGCTCAGCAGACTCGAGGTGAGGAGGTGCCAAGAAACCTCCAGACCGCTGCTGAC 120
XX
XX 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
XX 121 AAAAACCCAGGTTCAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
XX
XX 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
XX 181 ACCTCCATCAACCGGTGTCGTGGACCGTTTACCACGGTGGTGGTACCGTACCATCGCT 240
XX
XX 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
XX 241 TCCCGAAGAGGTCCGTTACCCAGATGTACACCAAGCTTGACAAAGACCTGGTGGTGG 300
XX
XX 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
XX

```

```

Db 301 CAGGCTCCCGAGGGTTCCTCCGTCCGCTGACCCCGTGCACCTCGGGTTCCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGGTTACCGCTCAGCTGACGTTATCCCGGTTCGTCGTCGGTGAAGTCCGCTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
Db 421 CTGCTGTCCCGGTCGATCTCCCTACCTGAAGGTTCCTCCGGTGGTCCGCTGCTGTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaIleValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCAGCTGTTGGTATCTTCGTCGTGCTGTTTCCACCGGTGGTGTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGGTGAAATCCCTGGAACCCACCATCGTTCCTCCCG 591
XX
XX RESULT 2
XX AAA73333
XX ID AAA73333 standard; DNA: 594 BP.
XX AC AAA73333;
XX
XX 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; muten; ds.
XX
XX Hepatitis C virus.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 1..594
XX /*tag= a
XX /product= "NS4A-NS3 fusion protein #6"
XX
XX WO2000040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI; 2000-465976/40.
XX DR P-PSDB; AAB15224.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26; Fig 16; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX

```



XX SQ Sequence 594 BP; 104 A; 101 C; 152 G; 147 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.8e-85 Length: 594  
 Score: 1005.00 Matches: 194  
 Percent Similarity: 99.49% Conservative: 2  
 Best Local Similarity: 98.48% Mismatches: 1  
 Query Match: 98.82% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73331 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 Db 1 ATGAAAAAAGAGGATCGGTGTTATCGCGCGGTATCAACCTGTCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40  
 Db 61 TACGCTCAGCAGACTCGAGGTGAGCAGGGTTGCCAGAGACCTCCACACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 Db 121 AAAAACACAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACTTCCTGGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 Db 181 ACCTCCATCAACGGTGTCTGTCGACCGTTTACACCGTGTGTTACCCGCTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 Db 241 TCCCGAAGGTCGGTGTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 Db 301 CAGGCTCCGAGGGTTCGGTTCCTGACCCGCTGACCTCGGTTCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 Db 361 CTGTTACCGGTACCGTACGCTGAGCTTATCCCGTTCGTCGTGTCGTCGTCGTCGTC 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 Db 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGAGTTCCTCCGTTGTCCTGCTGTGC 480  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 Db 481 CCGGTGTGTCACGCTGTTGGTATCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 Db 541 GCTGTTGACTTCATCCCGTGAATCTCCCTGGAACCCACCATCGCTTCCCG 591

RESULT 3

AAA73331  
 ID AAA73331 standard: DNA; 594 BP.

XX AC

XX AC

XX AC

XX 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein; ds.

XX Hepatitis C virus.

OS Synthetic.

XX Key

PH Location/Qualifiers

FT CDS 1..594

FT /\*tag= a

FT /product= "NS4A-NS3 fusion protein #4"

XX PN WO200040707-A1.  
 XX 13-JUL-2000.  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V.  
 DR WPI: 2000-465976/40.  
 DR P-PSDB; AAB15222.  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX Claim 26: Fig 14; 66pp; English.  
 PS The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0-1 variant.  
 XX SQ Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 3.44e-85 Length: 594  
 Score: 1002.00 Matches: 194  
 Percent Similarity: 98.48% Conservative: 0  
 Best Local Similarity: 98.48% Mismatches: 3  
 Query Match: 98.53% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73331 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 Db 1 ATGAAAAAAGGATCGGTGTTATCGTCGCGCGGTATCAACCTGTCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40  
 Db 61 TACGCTCAGCAGACTCGAGGTGAGGAGGTTGCCAAGAAACCTCCACACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 Db 121 AAAAACACAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACTTCCTGGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 Db 181 ACCTGCATCAACGGTGTCTGTCGACCGCTTACACCGGTGTCGTCGTCGTCGTCGTCGTC 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 Db 241 TCCCGAAGGTCGGTGTACCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 Db 301 CAGGCTCCGAGGGTTCGGTTCCTGACCCGCTGACCTCGGTTCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140

```

|||||
361 CTGTTACCCGTCACGTCGACGTTATCCCGTTCGTCGTCGACGCCGTTGCTCC 420
|||||
141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
|||||
421 CTGCTGTCCCGCGCTCCGATCTCTAGCTGAAAGGTTCTCTCCGTTGGTCCGCTGTGTC 480
|||||
161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
|||||
481 CCGCTGTGTCACGCTGTGGTATCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTA 540
|||||
181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
541 GCTGTTGACTTCACTCCCGGTTGAATCCCTGAAACACCATGCTGCCCG 591

RESULT 4
AAA73334
ID AAA73334 standard; DNA; 594 BP.
AC AAA73334;
DT 19-DEC-2000 (first entry)
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX Hepatitis C virus.
OS Hepatitis C virus.
OS Synthetic.
XX Key Location/Qualifiers
FT CDS 1..594
FT /*tag= a
FT /product= "NS4A-NS3 fusion protein #7"
XX
XX WO200040707-A1.
XX
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
XX
XX 08-JAN-1999; 99US-0115271.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX WittekInd M, Weinheimer S, Zhang Y, Goldfarb V;
XX
XX WPI: 2000-465976/40.
XX
XX P-PSDB; AAB15225.
XX
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX
XX Claim 26; Fig 17; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polyprotein produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-7 variant.
XX
XX Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

```

```

Alignment Scores:
Pred. No.: 1,56e-84 Length: 594
Score: 995.00 Matches: 193
Percent Similarity: 98.98% Conservative: 2
Best Local Similarity: 97.97% Mismatches: 2
Query Match: 97.84% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73334 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCCGTTGTTATCGTCGGCGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTGAGCGGTACCCGAGAGACCTCCACACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValIleGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCCAGGTGAAGGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGTGTCTGTGACCGGTTTACACCGGTGCTGACCGGTACCCGTCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGAAAGGTCGCGTTACCCAGATGTACACCAACGTTGACAAAGACCTGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCGCGAGGTTCCGTTCCCTGACCCCGTGCACCTCGCGGTCTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGCTCAGCTGAGCTGATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGCTCCGATCTCTACCTACCTGAAAGGTTCTCCGCTGCTGCTGCTGCTG 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGTGTCAGCTGTGGTATCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCTGCCGTCGTCGTCGTCG 591

RESULT 5
AAA73330
ID AAA73330 standard; DNA; 594 BP.
XX
XX AAA73330;
XX
XX 19-DEC-2000 (first entry)
XX
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.
XX
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT CDS 1..594
FT /*tag= a
FT /product= "NS4A-NS3 fusion protein #3"
XX
XX WO200040707-A1.
XX

```

```

PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
XX
DR P-PSDB; AAB15221.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
PS Claim 26; Fig 13; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;

Alignment Scores:
Pred. No.: 1.36e-83 Length: 594
Score: 985.00 Matches: 191
Percent Similarity: 96.95% Conservative: 0
Best Local Similarity: 96.95% Mismatches: 6
Query Match: 96.85% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73330 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThraLa 20
DB 1 ATGAAAAAAGAGATCGCTTGTATCGTCGGCGGTATCAACCTGCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40
DB 61 TAGCGTCAACGACTCGAGGTGAGGAGGTTGCCAAGAAACCTCCAGACCGGTGAC 120
QY 41 LysAsnGlnValGluGlyGlnValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAACCAGTTGAGGTTGAGATTCAGATCGTTCCACCGGTGCTCAGACCTTCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGTGTTCGTGGACCGTTTACCACGGTGTGTTACCGGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCCGAAAGGTCGGGTTCATCCAGATGTACACCAACGTTGACAAAGACTCTGGTTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CCGGCTCCGAGGGTTCCTCCCTCCGACCGGTGCTGCGGTTCCTCCGACCTGTGAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGGTACACGCTACGCTTATCCGGTTCGTCGTGTTGCTGCTCCCGTGGTTC 420

```

---

```

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGGTCCGATCTCTACCTGAAGAGTTCCTCCGCGTGTGCTGTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CGGCTGTGTCAGCGTGTGGTATCTTCGCTGCTGTTTGACCCCGGTGTGCTAAA 540
QY 181 AlaValAspPheProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 CGTGTGACTTCATCCCGTTCAATCCCTGGAACCAACCATGCTGCCCG 591
RESULT 6
AAA73329
ID AAA73329 standard; DNA; 588 BP.
XX
AC AAA73329;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutain; ds.
XX
OS Hepatitis C virus.
XX
XX Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..588
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #2"
XX
PN WO200040707-A1.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00345.
XX
PR 08-JAN-1999; 99US-0115271.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX
DR WPI; 2000-465976/40.
XX
DR P-PSDB; AAB15220.
XX
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT amino acid, useful for screening inhibitors that may treat hepatitis C
PT
XX
SQ Claim 26; Fig 12; 66pp; English.
XX
CC The present sequence is the coding sequence for a mutated version of a
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC protease enzymes. These proteins are both essential for the replication
CC of the virus, acting to cleave its replicative proteins from the
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins
CC should be effective as antiviral treatments of HCV infection. This is
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC failure and liver cancer. The present invention concerns a number of NS3
CC mutants and NS3-NS4A fusion proteins which can be used to identify
CC inhibitors of this type, as well as enabling structural studies of the
CC protease and protease-inhibitor complexes. The protein produced from this
CC sequence contains the alpha-helix0-1 variant.
XX
SQ Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

Alignment Scores:
Pred. No.: 2.1e-80 Length: 588
Score: 951.00 Matches: 187

```

Percent Similarity: 95.43% Conservative: 1  
 Best Local Similarity: 94.92% Mismatches: 7  
 Query Match: 93.51% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-18 (1-197) x AAA73329 (1-588)

QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAATAAAGGATCCGTTTATCGTCGCCGTATAGTACTGAACGGT-----GCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40  
 DB 55 TAGGCTCAGCAGCTCGAGGTGAGGAGGTTCGCAAGAACCTCCACACCGGTGCTGAC 114  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 115 AAAAACCGAGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCACACCTTCCTGGCT 174  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 175 ACCGTGATCAACGGTGTTCGTCGACCGTTTACCACGGTGTGTTACCGTACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 DB 235 TCCCGGAAAGGTCCGGTTATCCAGATGTACACCAAGCTTGACAAAGACCTGGTTGGTGG 294  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 295 CCGGCTCCGAGGGTTCGCTCCGTCACCGCTGACCGTGGTTCCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 355 CTGGTTACCGTCACGCTGACGTTATCCGGTTCGTCGTCGTGACTCCCGTGGTTC 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 415 CTGCTGTCCCGGTCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCTGCTGTGC 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 475 CCGGCTGTCCACGCTGTGTGTATCTTCCTGCTGCTGCTGTTTGCACCGCTGTGCTAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCAATCCCGTGTGAATCCCTGGAACACCATCGCTTCCCG 585

## RESULT 7

AAA73335  
 ID AAA73335 standard; DNA; 594 BP.

XX AC AAA73335;

XX DT 19-DEC-2000 (first entry)

XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.

XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 liver failure; liver cancer; mutant; mutein; ds.

XX OS Hepatitis C virus.

XX OS Synthetic.

XX FH Key Location/Qualifiers  
 FT CDS 1..594

FT /tag= a  
 /product= "NS4A-NS3 fusion protein #8"

PN WO200040707-A1.

XX 13-JUL-2000.

XX PD 06-JAN-2000; 2000WO-US00345.

PF

XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15226.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT .  
 XX Disclosure; Fig 18; 66pp; English.  
 XX The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0 wild-type sequence.  
 XX Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;  
 SQ

## Alignment Scores:

Pred. No.: 6.27e-80 Length: 594  
 Score: 946.00 Matches: 186  
 Percent Similarity: 94.42% Conservative: 0  
 Best Local Similarity: 94.42% Mismatches: 11  
 Query Match: 93.02% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-18 (1-197) x AAA73335 (1-594)

QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAATAAAGGATCCGTTTATCGTCGCCGTATAGTACTGAACCTGTCCGGTACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAsp 40  
 DB 61 TAGGCTCAGCAGCTCGAGGTGAGGAGGTTCGCAAGAACCTCCACACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAAACCGAGTTGAAGGTGAAGTTCAGATCGTTTCCACCGCTGCACACCTTCCTGGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCGTGATCAACGGTGTTCGTCGACCGTTTACCACGGTGTGTTACCGTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 DB 241 TCCCGGAAAGGTCCGGTTATCCAGATGTACACCAAGCTTGACAAAGACCTGGTTGGTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CCGGCTCCGAGGGTTCGCTCCGTCACCGCTGACCGTGGTTCCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGGTTACCGTCACGCTGACGTTATCCCGGTTTCGTCGTCGTCGTCGTCGTCGTC 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGGTCGATCTCTACCTGAAGGTTCCTCCGGTGGTCCGCTGCTGTGTC 480

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 Db 481 CCGGCTGGTCACGCTGTGGTATCTCCGCGCGCTGTGGTACCGGCTGTGCTAA 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 Db 541 GCTGTTGACTTCATCCCGGTGAATCCCTGGAACACCATCGGTTCCCGG 591

RESULT 8  
 AAA73328  
 ID AAA73328 standard: DNA: 588 BP.  
 XX  
 AC AAA73328;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..588  
 FT /tag- a  
 FT /product- "NS3-NS4A fusion protein"  
 XX  
 FN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 990S-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI: 2000-465976/40.  
 DR P-PSDB; AAB15212.  
 XX  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT .  
 XX  
 PS Disclosure; Fig 10; 66pp; English.  
 XX  
 CC The present sequence is the coding sequence for a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.  
 XX  
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.68e-77 Length: 588  
 Score: 912.00 Matches: 182  
 Percent Similarity: 92.89% Conservative: 1  
 Best Local Similarity: 92.39% Mismatches: 12  
 Query Match: 89.68% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-18 (1-197) x AAA73328 (1-588)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 Db 1 ATGAATAAAAGAGTTCGGTTGTTATCGTCGCCGTATAGTACTAGAACGGT-----GCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlnThrSerGlnThrGlyArgAsp 40  
 Db 55 TACGCTACGACACTCGAGGTCTCGTGGTTGTCATCATCCTCCCTGACCGGTCTCGTGC 114  
 QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 Db 115 AAAAACCCAGGTGAAGGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTTCCTGGCT 174  
 QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 Db 175 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTGTACCGGTACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 Db 235 TCCCGAAGGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 294  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 Db 295 CCGGCTCCGACAGGTTCCTCGTACCCCGCTGCACCTCGCGGTTCCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 Db 355 CTGGTTACCCGCTACGCTGACGTTATCCCGGTTCGTGCTGCTGCTGCTGCTGCTGCTGCT 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 Db 415 CTGCTGTCCTCCGCTCCGATCTCCTACCTGAAGGTTCTCCGCTGCTGCTGCTGCTGCTG 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 Db 475 CCGGCTGCTACGCTGTTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 Db 535 GCTGTTGACTTCATCCCGGTGAATCCCTGGAACACCATCGGTTCCCGG 585

RESULT 9  
 ABA95615  
 ID ABA95615 standard: DNA: 12734 BP.  
 XX  
 AC ABA95615;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Chimeric BVDV/HCV NS3-wt sequence.  
 XX  
 KW Pestivirus; Npro; protease; NS3; screening; ds.  
 OS Chimeric - Bovine viral diarrhea virus.  
 OS Chimeric - Hepatitis C virus.  
 XX  
 PN US6326137-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 PF 25-JUN-1999; 990S-0344456.  
 XX  
 PR 25-JUN-1999; 990S-0344456.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Hong 2, Lai VCH, Lau JYN;  
 XX  
 DR WPI; 2002-121103/16.  
 XX  
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3  
 PT protease gene, useful for in vivo screening of compounds which inhibit  
 PT HCV infection

XX Example 2: Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a  
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
 CC each junction site recognised by the Npro protease is replaced with a  
 CC junction site recognised by the HCV NS3 protease. The construct is useful  
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
 CC protease, where screening may be in cell culture or in an animal model.  
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.06e-73 Length: 12734  
 Score: 892.50 Matches: 177  
 Percent Similarity: 92.82% Conservative: 4  
 Best Local Similarity: 90.77% Mismatches: 11  
 Query Match: 87.76% Indels: 3  
 DB: 24 Gaps: 1

US-09-965-594-18 (1-197) x ABA95615 (1-12734)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThralaTyr 21  
 Db 413 GGTAGTGTGTTTATTTGTTAGTATGTTTATCTGTTAGTGGTAGTATCACGGCGTAC 472  
 Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 Db 473 GCCCAGCAGAGAGGCGCTCTAGGCTGTAAGATCACCGACTGCTGCGCGGACAAA 532  
 Qy 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 Db 533 AACCAAGTGGAGGTGAGGTGCCAGATCGTGTCACTGCTACCAACCTTCTCTGCAACG 592  
 Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 Db 593 TGCATCAATGGGTATGCTGGAGCTGTCTACCGGGCGGCGAACGAGCATCGCATCA 652  
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 Db 653 CCCAAGGGTCTCTCATCTCAGATGTATACCAATGTGGACCAAGACCTTGTGGCGTGC 712  
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 Db 713 GCTCTCTCAAGTTCCTCGCTCATTCACACCTCGACCTGCGGCTCTCGGACCTTTACCTG 772  
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 Db 773 GTTACGAGGCACCGCCGCTCATTCCTCGTGGCGGCGAGGTGATAGCAGGGGTAGCCTG 832  
 Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerCysGlyProLeuLeuCysPro 161  
 Db 833 CTTTCGCGCGCGGCCATTTCTTACCTAAAGGCTCTCTCGGGGGGTCCTGTGTGTGCCCC 892  
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 Db 893 CGGGACACCCGCTGGGCGCTATTTCAGGGCGCGGCTGTGCACCCGTGGAGTGCCCAAGCG 952  
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 Db 953 GTGACATTATCCCTGTGGAGACCTTAGAGACAAACCATGAGATCC 997

RESULT 10

AAx80355

ID AAx80355 standard; cDNA; 1998 BP.

XX

AC AAX80355;  
 XX 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX WO9928482-A2.  
 XX 10-JUN-1999.  
 XX 24-NOV-1998; 98WO-US24528.  
 XX 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;  
 PI WPI; 1999-385385/32.  
 DR New hepatitis C virus covalent complexes  
 PT Disclosure; Page 166-169; 21lpp; English.  
 XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

XX Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.27e-73 Length: 1998  
 Score: 881.50 Matches: 167  
 Percent Similarity: 93.37% Conservative: 16  
 Best Local Similarity: 85.20% Mismatches: 10  
 Query Match: 86.68% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAX80355 (1-1998)

Qy 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThralaTyr 21  
 Db 64 GGTCTCTGTTTATTTGTTAGTATGTTTATCTGTTAGTGGTAGTATCACGGCGTAC 123  
 Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 Db 124 TCCCAACAGACGGCGGCGCTCTAGGCTGTCAGAGACTAGCTTACAGCGCGGACAG 183  
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 Db 184 AACCAAGTTCGAGGAGAGGTTCAGGTGGTTTCCACCGCAACACATCTCTCTCGCGACC 243  
 Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 Db 244 TGCGTCAACGGCGTGTGTTGGACCTTTACCATGTTGCTGCTCAAGACCTTAGCCGCG 303

QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGCCCAATACCCAGATGTACATAATGTGGACCAGGACCTGTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 CGCCCCCGGGCGGCTCTTGTACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgValAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATCTGACGTCATTCGGGTGCGCGCGGGGGGACAGTAGGGGGGCGCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGAGGCTGTCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGCACGCTGTGGGCATCTTCCGGGTGCGGTATGACCGCGGGGGTTCGGAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 11  
 AAX80359  
 ID AAX80359 standard; cDNA; 1998 BP.  
 XX  
 AC AAX80359;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX  
 DR WPI; 1999-385385/32.  
 XX  
 PT New hepatitis C virus covalent complexes  
 XX  
 PS Disclosure; Page 179-182; 21pp; English.  
 XX  
 CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 6.26e-73 Length: 1998  
 Score: 878.50 Matches: 166  
 Percent Similarity: 93.37% Conservative: 17  
 Best Local Similarity: 84.69% Mismatches: 10  
 Query Match: 86.38% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAX80359 (1-1998)  
 QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTCTCTGTTATGTTGTTGTTAGTAATATTTATCTGGTAGGTAGTATACGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACGGCGGCTACTTGGTTGCAAGAAGACTAGCCTTACAGCCGGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACAGGTCCAGGAGAGGTTTCAGGTGTTTCCACCGCAACATCTCTCTGCGGACC 243  
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TGCCTCAACGGCGTGTGTGGACCGTTTACCATGGTCTGGCTCAAAGACCTTAGCGGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGGGCAATCACCAGATGTACACTAATGTGGACCAGACCTCGTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 CGCCCCCGGGCGGCTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGTCATTCGGGTGCGCGCGGGGGGACAGTAGGGGAGCCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCysPro 161  
 DB 484 CTCCTCCCGAGGCTGTCTCTACTTGAAGGGTCTCTGTTGTTGTTGTTGTTGTTGTTG 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGCACGCTGTGGGCATCTTCCGGGTGCGGTATGACCGCGGGGGTTCGGAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 12  
 AAX80353  
 ID AAX80353 standard; cDNA; 1998 BP.  
 XX  
 AC AAX80353;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.





```

Db 124 TCCCAACAGACGGGGCTTACTTGGTTCATCAAGACTAGCCCTTACAGCGCGGACAG 193
Qy 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCAGGAGAGAGTTCAGTGGTTCCACCGCAACACAAATCTTCTCGGCGACC 243
Qy 62 SerIleAsnGlyValLeuIleThrValThrHisGlyValAlaGlyThrArgThrIleAlaSer 81
Db 244 TGGGTCAACGGCGTGTGGACCGTTTACATGGTGTGCTGCTCAAGACCTTAGCGGCG 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACCGACCTCGTGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GCGCCCCCGGGGGCGGTCTCTTGACACCATGACCTGTGGCGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGAGACATGCTGACGTCAITCCGGTGGCGGGGGCGGACAGTAGGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuGlyGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTTCGGGTGGTCCACTGCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValAlaGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGACCGCTGTGGCGATCTCCGGGTGCGGTATGCACCCGGGGGGTTCGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACTTTGGCCCGTAGAGTCCATGGAACACTACTATGCGGTCTCG 651

```

## RESULT 14

ABX15706  
ID ABX15706 standard; DNA; 612 BP.

XX AC ABX15706;

XX DT 28-MAR-2003 (first entry)

XX DE Anti-viral synthetic prototoxophore associated DNA sequence.

XX KW Hepatitis C; ds; viral prototoxophore; anti-viral; tumour;  
XX KW virus; infection; antitumour; toxophore; human immunodeficiency virus;  
XX KW HIV infection; herpes simplex virus; HSV; rhinovirus; NS3 protease.

XX OS Unidentified.

XX PN WO200287500-A2.

XX PD 07-NOV-2002.

XX PF 26-APR-2002; 2002WO-US13223.

XX PR 27-APR-2001; 2001US-286893P.

XX XX (NEWB-) NEWBIOTICS INC.

XX PI Cathers BE, Neuteboom STC, Shepard HM;

XX XX WPI; 2003-167102/16.

XX XX Novel synthetic viral prototoxophore for treating viral infections, has  
PT toxin moiety incorporated into substrate domain specific for viral  
PT enzyme, bound and modified by viral enzyme to get converted into  
PT toxophore

XX PS Example 1; Page 62; 66pp; English.

XX XX This invention relates to a novel synthetic viral prototoxophore

comprising a toxin moiety operatively incorporated into a substrate domain specific for a viral enzyme. This prototoxophore may be bound and modified by the viral enzyme thus converting it to a toxophore. Also disclosed in the invention is a method for enhancing the anti-viral effect of an antiviral agent, this method comprises contacting a cell, infected with a virus or is susceptible to infection, with a prototoxophore. The invention further comprises an assay to identify anti-viral agents, comprising contacting an infected cell with a candidate agent and comparing the ability of the agent to inhibit the growth or infectivity of the virus in the cell. The prototoxophores of the invention may have virucide or antitumour activity. The prototoxophores of the invention may be useful for reducing or inhibiting viral infectivity, by contacting a cell (e.g. lymphocyte, nerve cell, connective tissue cell, muscle cell or hepatocyte) which is infected with a virus or is susceptible to infection with a virus, with an effective amount of the prototoxophore. The cells are cell lines adapted to long term continuous culture or isolated from a subject. The prototoxophore is also useful for ameliorating the severity of a viral infection in a subject, where the virus is selected from human immunodeficiency virus (HIV), herpes simplex virus (HSV), rhinovirus and hepatitis virus, by administering an effective amount of the prototoxophore to the subject. The prototoxophores of the invention are also useful for treating tumours. The present sequence represents an antiviral prototoxophore associated DNA sequence, this sequence is described as a recombinant NS3/NS4 fusion protein in example 1 of the invention although it is clearly not a protein sequence.

XX SQ Sequence 612 BP; 120 A; 171 C; 191 G; 130 T; 0 other;

## Alignment Scores:

Pred. No.: 3,39e-73 Length: 612  
Score: 874.50 Matches: 175  
Percent Similarity: 91.28% Conservative: 3  
Best Local Similarity: 89.74% Mismatches: 14  
Query Match: 85.99% Indels: 3  
DB: 25 Gaps: 1

US-09-965-594-18 (1-197) x ABX15706 (1-612)

```

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GGTAGTGGTCTATTGGGTAGGATCATTTTGTCCGGTAGTGGTAGTATCATCGCGTAC 78
Qy 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSerGlnThrGlyArgAspLys 41
Db 79 GCCACGACACAAAGGGCGCTCTAGGGTGCATAATCACCGCTAACTGGCGGGACAAA 138
Qy 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTGGAGGTGAGGTCCAGATTGTGTCACCTGCTGCCCAACCTTCTTGGCAACG 198
Qy 62 SerIleAsnGlyValLeuIleThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTGCTGGACTGCTACACGGGGCGGACGACGACCATCGCTCA 258
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 259 CCCAAGGTCTCTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCC 318
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGAAGGTACCCGCTCATTCACCTTGGCGGTCTCTCGGACCTTTACCTG 378
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGCACCGCGATGTCTATCCCGTCCCGCGGGGTGATAGCAGGGGACGCTG 438
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGTCGCCCGCGCCATTTCCTACTTGAAGGCTCTCGGGGGTCCGCTGTGTGGCCCC 498
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

```

Db 499 GCGGGCAGCCGTCGGCATATTTAGGCGCGGTGTCACCCGTGGAGTGGCTAAGGCG 558

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 |||||  
 Db 559 GTGAGCTTTATCCCTGTGGAGAACCTAGAGACACCATGAGGTCC 603

RESULT 15  
 AAX80345  
 ID AAX80345 standard; cDNA; 651 BP.  
 AC AAX80345;  
 XX 07-SEP-1999 (first entry)  
 DT  
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 OS  
 XX WO9928482-A2.  
 PN 10-JUN-1999.  
 XX 24-NOV-1998; 98WO-US24528.  
 PF 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 PA  
 XX Malcolm BA, Taremi SS, Weber PC, Yao N;  
 PI WPI; 1999-385385/32.  
 DR  
 XX New hepatitis C virus covalent complexes  
 PT  
 XX Disclosure: Page 147-148; 21pp; English.  
 PS  
 XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.  
 XX  
 SQ Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.66e-73 Length: 651  
 Score: 874.50 Matches: 166  
 Percent Similarity: 93.33% Conservative: 16  
 Best Local Similarity: 85.13% Mismatches: 10  
 Query Match: 85.99% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-18 (1-197) x AAX80345 (1-651)

Qy 5 GlySerValIleValGlyArgGluAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTTCCTGTGTTATGTGTGGTAGAATATTATTCGTGGTAGTGGTAGTACGGGCTAC 123  
 |||||

Qy 22 AlaGlnGlnThrArgGlyGluGluGlyCysGlnGluThrSerGlnThrGlyArgAspLys 41  
 ::::|

Db 124 TCCCAACACAGCGGGGCGCTACTTGGTTCGAAGAAGACTAGCTTTACAGCGGGGACAAG 183

Qy 42 AsnGlnValGlnGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 AACCAAGTCCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACACATCTCTCTGGCGACC 243  
 ::::|

Qy 62 SerIleAsnGlyValLeuIleThrValTyrHisGlyValGlyValThrArgThrIleAlaSer 81  
 |||||  
 Db 244 TCGGTCAACGGCGGTGTGTGGACCGTTTACCATGGTGTGGCTCAAGACCTTAGCCGGC 303  
 |||||

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 |||||  
 Db 304 CCAAGGGGCAATACACCCAGATGTACATAATGTGGACACAGGACCTCGTGGCTGGCAG 363  
 |||||

Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||

Db 364 GCGCCCCCGGGGCGGCTTCTTGACACCTGACCTGTGGCAGCTCAGACCTTTACTTG 423  
 |||||

Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 |||||

Db 424 GTCACGAGACATGCTCAGCGTCATTCGGTGCCTGGCGGGGGGACAGTAGGGGAGCCTG 483  
 |||||

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 |||||

Db 484 CTCTCCCCCAGGCGCTGCTCCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543  
 |||||

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 ::::|

Db 544 TCGGGCACGCTGTGGGCATCTTCGGGCTGCCGTATGCACCCGGGGGTTGCGAAGGCG 603  
 |||||

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 |||||

Db 604 GTGGACTTTGTGCGCGCTAGAGTCCATGGAACACTACTATGCGGTCT 648

Search completed: August 30, 2003, 19:48:09

Job time : 188.939 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-18  
Perfect score: 1017  
Sequence: 1 MKKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETMRSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip  
-O=/cgn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151919\_28322/app\_query.fasta\_1.2872  
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGCN\_1\_12630 &runat\_29082003\_151919\_28322 -NCP0=6 -ICPU=3  
-NO\_WMAP -LARGESQUEERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=170 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	106	10.4	1146	12	BM915803 AGENCOURT
C 2	104.5	10.3	1031	14	CB950999 AGENCOURT
C 3	102.5	10.1	1403	13	BQ926101 AGENCOURT
C 4	100.5	9.9	1199	13	BQ992487 AGENCOURT
C 5	95.5	9.4	502	9	AA036834 2K29405.r
C 6	95.5	9.4	701	10	BF863244 963042C02
C 7	95.5	9.4	984	10	BF304699 601888252
C 8	95	9.3	629	10	BG089727 mad90e06
C 9	93.5	9.2	905	13	BUS42842 AGENCOURT
C 10	93.5	9.2	1733	12	BM553374 AGENCOURT
C 11	93	9.1	528	12	BM402566 SLA005F12
C 12	93	9.1	644	29	BX238988 Danio rer
C 13	93	9.1	701	29	BZ342381 ic83b11.b
C 14	93	9.1	1213	13	BUS41777 AGENCOURT
C 15	92.5	9.1	772	29	CC406704
C 16	92.5	9.1	789	29	CC406705
C 17	92.5	9.1	938	13	BQ994657 AGENCOURT
C 18	91.5	9.0	528	28	AQ620249 HS_5182.B
C 19	91.5	9.0	1294	13	BQ925457 AGENCOURT
C 20	91	8.9	580	14	CA728398 wd11c.pk0
C 21	91	8.9	646	12	BG853999 1024038G0
C 22	91	8.9	753	13	BD402910 604139183
C 23	91	8.9	865	13	BD219343 603758452
C 24	91	8.9	906	13	BX34207 BX434207
C 25	90.5	8.9	814	11	CNS09179
C 26	90.5	8.9	1141	11	AK080545
C 27	90.5	8.9	1440	12	BM467279
C 28	90	8.8	500	12	BM708007
C 29	90	8.8	569	12	BM825317
C 30	90	8.8	617	10	BE055938
C 31	90	8.8	622	9	AV835401
C 32	90	8.8	631	10	AW961059
C 33	90	8.8	658	12	BM830847
C 34	90	8.8	757	12	B1258851
C 35	90	8.8	763	13	BX093694
C 36	90	8.8	812	13	BO434921
C 37	90	8.8	859	13	BQ222616
C 38	90	8.8	987	13	BX357170
C 39	90	8.8	1015	13	BX404631
C 40	90	8.8	1049	29	CNS040AW
C 41	90	8.8	1123	13	BX398349
C 42	90	8.8	1201	9	AL562877
C 43	90	8.8	1201	13	BX339469
C 44	90	8.8	1564	12	BG827880
C 45	89.5	8.8	409	14	CB805033

# ALIGNMENTS

RESULT 1  
BM915803/C  
LOCUS BM915803 1146 bp mRNA linear EST 12-MAR-2002  
DEFINITION AGENCOURT\_6639455 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5482056  
5', mRNA sequence.  
ACCESSION BM915803  
VERSION BM915803.1 GI:19366182  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1146)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LCM2007 row: i column: 01  
 High quality sequence start: 6  
 High quality sequence stop: 256.  
 High quality sequence stop: 256.  
 Location/Qualifiers  
 1..1146  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5482056"  
 /tissue\_type="amelanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH-MGC-41"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."  
 BASE COUNT 169 a 492 c 344 g 141 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,1 Length: 1146  
 Score: 106.00 Matches: 47  
 Percent Similarity: 36.26% Conservative: 19  
 Best Local Similarity: 25.82% Mismatches: 59  
 Query Match: 10.42% Indels: 57  
 DB: 12 Gaps: 10  
 US-09-965-594-18 (1-197) x BM915803 (1-1146)  
 QY 22 AlaGlnGlnThrArgGlyGluGlyCysGlnGlnThrSer-GlnThrGlyArgAspLy 41  
 Db 1098 GCGCAGACGGTGTGGCGAGCGAGGGTGTCTCCGCCCTCGTACCGTCCGTGGAG 1039  
 QY 41 sAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaTh 61  
 Db 1038 CTGGAGGACAGAGT-----CTACGCGGTGGGTAGGGGA 1003  
 QY 61 rSerIleAsnGlyValLeuThrThrValThrHis-----GlyAlaGlyThr-- 76  
 Db 1002 CGCCCTGTGGTGTGTGGT-----TATCACTTCCGCGCGCGGGGGAGTACGTG 949  
 QY 77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90  
 Db 948 AGCGAGGGGCGCGCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903  
 QY 90 rThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGln-----GlySerAr 107  
 Db 902 -----CAGATGTGGGTGGAGAGCCGCCCTCGCGGGTGGGGGCCAG 859  
 QY 107 gSerLeuThrProCysThyGlySerSerAspLeuThrValThrArgHisAlaAs 127  
 Db 858 ACTTCTGTGTCTGTTCTGTGG----- 834  
 QY 127 pValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIl 147  
 Db 833 -----CGGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 787

QY 147 eSerTyrlLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValAlG1 167  
 Db 786 CGGTATCTACAGCGCGCGACGACACATCTCTCTCGG-----TG 742  
 QY 167 yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IlePr 186  
 Db 741 GGCCTTCCGGCTGTGTGTCTCTCGCGTCTCCGCGGGGGGGGGTTCGCGTACC 682  
 QY 186 oVal 187  
 Db 681 TTGTG 678  
 RESULT 2  
 CB950999  
 LOCUS 1031 bp mRNA linear EST 29-APR-2003  
 DEFINITION AGENCOURT\_13445496 NIH-MGC\_177 Mus musculus cDNA clone  
 IMAGE:30316162 5', mRNA sequence.  
 ACCESSION CB950999  
 VERSION CB950999.1 GI:30205777  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDCM107 row: b column: 11  
 High quality sequence stop: 333.  
 High quality sequence stop: 333.  
 Location/Qualifiers  
 1..1031  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30316162"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH-MGC\_177"  
 /note="Organ: liver; Vector: pDNR-LIB; Site\_1: SfiI (ggccattatggcc); Site\_2: SfiI (ggccctcgcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCAACGAGTGGCCATTCAGCGCGG-3' and  
 5'-ATTCAGAGCGGAGCGGCGGACATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH-MGC Library."  
 BASE COUNT 235 a 309 c 211 g 275 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,8 Length: 1031  
 Score: 104.50 Matches: 51  
 Percent Similarity: 41.10% Conservative: 16  
 Best Local Similarity: 31.29% Mismatches: 62  
 Query Match: 10.28% Indels: 35  
 DB: 14 Gaps: 8  
 US-09-965-594-18 (1-197) x CB950999 (1-1031)  
 QY 44 ValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerile 63  
 Db 395 ATTCAGGCTATCCCAACAAAGAGAGTACATCGCAGCTTTTCTCTT---CACTCAIT 451

Qy 64 AsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83  
 Db 452 TTGGGCACACTGTGTCGGTGGACAT-----ATCGATCCCTTAA 493

Qy 84 GlyProValThrGlnMetTyrThrAsnValAspLysValGlyTyrGlnAlaPro 103  
 Db 494 GGGCCTTTACAAAA-----ACACTTAACCT-CCTTGCCTGGCTGGCATGTGG 543

Qy 104 Gln-----GlySerArgSerLeuThrProCysThrCysGlySerSerAsp 118  
 Db 544 CAAAGAACCGTTTTGGGTTCGGCTCTTGGCCCCCCCCCAATTTGGAACCACTGGC 603

Qy 119 LeuTyrLeuValThrArgHisAlaAsp-ValIleProValArgArgGlyAspSerAr 138  
 Db 604 -----ACCACCATGGCGCTGTGTTCCTGGCTCTCCGCTGGCAATAC 651

Qy 138 gGlySerLeuLeuSerProArgProfileSerTyrLeuLysGlySerSergly----- 155  
 Db 652 AAACNCCCTTAACCGTCCCTCCCAACAATATTCTTCAAGCGTCTCTGATTCCCTAA 711

Qy 156 -GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerTh 175  
 Db 712 GTCCCTTCCTTTTACCCGACACCATTTGTGGGACACAGCGCTCTTTTATCTTC 771

Qy 175 rArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetAr 195  
 Db 772 C-----CCCTCATGCTCTT---CCACACGGCG 798

Qy 195 gSerPro 197  
 Db 799 AACACCC 805

RESULT 3  
 BQ926101/C  
 LOCUS  
 DEFINITION AGNCOURT\_8752655 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6335718  
 5', mRNA sequence.

ACCESSION BQ926101  
 VERSION BQ926101.1 GI:22341132  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1403)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
 Ph.D.  
 cDNA Library Preparation: ResGen, Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM13798 row: j column: 07  
 High quality sequence stop: 101.  
 Location/Qualifiers  
 1..1403  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6335718"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_130"  
 /note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdbd;  
 Site 1: EcorV; Site 2: NotI; Cloned unidirectionally.  
 Primer: Oligo dT. Average insert size 1.95 kb.  
 Constructed by ResGen, Invitrogen Corp. Note: this is a

BASE COUNT 297 a 521 c 237 g 345 t 3 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8.97 Length: 1403  
 Score: 102.50 Matches: 59  
 Percent Similarity: 35.50% Conservative: 12  
 Best Local Similarity: 29.50% Mismatches: 67  
 Query Match: 10.08% Indels: 62  
 DB: 13 Gaps: 11

US-09-965-594-18 (1-197) x BQ926101 (1-1403)

Qy 11 GlyArgIleAsnLeuSerGlyAspThrAlaGlnGlnThrArgGlyGluGluGly 30  
 Db 1378 GGGTGTTCANCGGTTCAGGACAGGTGCGCC---GCACACTCGACGCTCGGCCAGAGACT 1322

Qy 31 CysGlnGluThr-----SerGlnThrGlyArgAspLysAsnGln-----Val 44  
 Db 1321 TGTGGGGCGCGGTGGCGCATACCCCGGTGGATCGAGTCCAGCGCGCTTGTATACA 1262

Qy 45 GluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsn 64  
 Db 1261 GAGGGGAAA-----CAG 1250

Qy 65 GlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys--G 84  
 Db 1249 GGGTA---TGTTATCAGCGGCTGGCGAGTACT-----TCCCTAAACGC 1205

Qy 84 LyrProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaPro 104  
 Db 1204 GCGCGTGGCGAGTATATACCGAGTGCAGGCGGCGGCGGCGTGAACGTTGACC 1145

Qy 104 InGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrA 124  
 Db 1144 AA---GAGAGCACCTGACGCCCTCCCTGTGGGCTGTGATATAACAAATGTGCG 1088

Qy 124 rHisAlaAspValIleProValArgArgGlyAsp----- 136  
 Db 1087 GGCACGGTGTGTGTACTACGCGGACGCGCTCCACGCGCTCTCTAACAGCGC 1028

Qy 137 -----SerArgGlySerLeuSerProArgProIle-SerTyrLeuLysGlySerSer 154  
 Db 1027 CCGCCTCCCGCGCACAC-----AGGTAATAATCATATCGCGCGGCGGATTTC 980

Qy 155 Gly-----GlyProLeuLeuCys 160  
 Db 979 GCATTCGCGGGAGAGCGCGGCTCGGGGGGCGCGCGCTGCGGCGCTGAGGCGC 920

Qy 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178  
 Db 919 AGGAGAGGC-----GGCGTGTTCGGCGGTGAGGACGAAAGCGCGCGTG 875

RESULT 4  
 BQ92487  
 LOCUS  
 DEFINITION AGNCOURT\_8417538 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
 IMAGE:6192708 5', mRNA sequence.

ACCESSION BQ92487  
 VERSION BQ92487.1 GI:22284501  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1199)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski

```

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13595 row: c column: 13
High quality sequence start: 57
High quality sequence stop: 394.
Location/Qualifiers
1. .ll199
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6192708"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_lib="Lupski_sympathetic_trunk"
/notes="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACGACGCGTCGC-3' and
5'-GACTAGCTTCTAGATCCGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."
255 a 362 c 343 g 211 t 28 others

```

871 TCAGGCGTTTAAAGCCCGCCGCTTCGGCGCGCGGAAGCA 913

Db

RESULT 5  
AA036834/c  
LOCUS  
DEFINITION  
502 bp mRNA linear EST 26-AUG-1996  
zk29dd05.r1 Soares\_pregnant\_uterus.MbHPU Homo sapiens cDNA clone  
IMAGE:471945\_5 , similar to PIR:A55195 A55195 chordin precursor -  
African clawed frog ;, mRNA sequence.

ACCESSION  
AA036834  
VERSION  
AA036834.1 GI:1509872  
KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens

REFERENCE  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS  
Hillier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M., Holman  
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marras,M., Parsons,J.,  
Rifkin,L., Rohlfing,I., Soares,M., Tan,F., Trevaskis,E., Waterston  
R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE  
The Washo-Merck EST Project

COMMENT  
Unpublished  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 300.

FEATURES	source	Location/Qualifiers	high quality sequence	stop: 300.
BASE COUNT	84 a	141 c	165 g	100 t
ORIGIN				
Alignment Scores:				
Pred. No.:	11, 6	Length:	502	
Score:	95.50	Matches:	37	
Percent Similarity:	38.51%	Conservative:	20	
Best Local Similarity:	25.00%	Mismatches:	45	
Query Match:	9.39%	Indels:	46	
DB:	9	Gaps:	6	
US-09-965-594-18 (1-197) x AA036834 (1-502)				
Qy	14	AsnLeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGluGlyCysGlnGlu	33	
Db	347	CATCTTGCAGGTGATACAGCTCATCTTCTCCAAAANGGGGGGCTAGGGGGTGCACAGCT	288	
Qy	34	ThrSerGlnThrGly-----ArgAspLysAsnGlnValGluGlyGluValGlnIleVal	51	
Db	287	CTGACTTCTTCTGGGAAACCACCTCCCGCAGCAACGCCCGGGGCCCATC-----	237	
US-09-965-594-18 (1-1199)				
Qy	68	TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThr	87	
Db	484	TGGGATCCATTTTAAATAAGGCTCTCTTAATCATGGCCGCCCGCTGATA	543	
Qy	88	GlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaProGlnLysArg	107	
Db	544	CTTCCATTATCCACATGTGACAGTGACTTT-----	573	
Qy	108	SerLeuThrProCysThr-----CysGlySerSerSerAsp	118	
Db	574	-----TGTTGGCTTCGCTCAGACAGCCGCCATGACCGATGTTGGGCTTATGGGAACGGCGAG	630	
Qy	119	LeuTyr-LeuValThr-----ArgHisAlaAspValIleProValArg-----	132	
Db	631	CGGTTTCATGGCCACTCCCTCCCTATAAAACAGCCCAACGCTGTTCCATGGGCGGGCT	690	
Qy	133	-----ArgArgGlyAspSerArgGlySerLeuLeu--	142	
Db	691	GGGTGTTTGGCAGGCGCAAGCGGGGTGGGGCATGGTAGGACTCGGGGGGCGATTCTCTG	750	
Qy	143	-----SerProArgProIleSerTyrLeuLys-----GlySerSerG1	155	
Db	751	AAACCCGACCTCGGCCCGCCACCGATGCGCTTAGCCCTCCCTTACAGCCACCGCCCGGG	810	
Qy	155	yGlyProLeuLeuLysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerTh	175	
Db	811	CCCCCCTTAACATCTCTTACCTCCCTGCGCGCGGGGGGAGACGTGGGGCCCATACGGGC	870	
Qy	175	rArgGlyValAlaLysAlaValAspPheIleProValGluSer	189	

```

Qy 52 SerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeu----- 67
Db 236 -----AGCTCGATGGGTGTCACACCAAAATTATCCCATACGGCCACTA 230
Qy 68 -----TTPThrValThrHisGlyAlaGlyThrArgThrIleAlaSerPro 82
Db 206 GGGCCCGACCCACCTGTCACAGTGTTCAGCAGTGGT----- 168
Qy 83 LysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTrpGlnAla 102
Db 167 -----GGGTTCACACGACACAGCTGGCA---CAGGCC 138
Qy 103 ProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuVal 122
Db 137 AGCGGGGACACCTTCACAGTGCACCTCCAGTGCCTCCAGTGCCTTCAGAGTGCAG 78
Qy 123 ThrArgHisAlaAspValIleProValArg-ArgArgGlyAspSerArgGlySerLeuLe 142
Db 77 ACAGCACACTTA-----ATTAGCCAAAGGGGGGCACAAAGGGGTGCCNANCN 30
Qy 142 uSerProArgProIleSerThr 149
Db 29 GTACCCGTTGCCGCCACGCTC 8

RESULT 6
BF863244
LOCUS
DEFINITION
963042C02.xl C. reinhardtii CC-1690, Stress condition I, normalized
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
BF863244
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
1 (bases 1 to 701)
REFERENCE
AUTHORS
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..701
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into Lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
173 a 213 c 175 g 140 t

```

```

Alignment Scores:
Pred. No.: 17.9 Length: 701
Score: 95.50 Matches: 38
Percent Similarity: 38.96% Conservative: 22
Best Local Similarity: 24.68% Mismatches: 63
Query Match: 9.39% Indels: 31
DB: 10 Gaps: 7

US-09-965-594-18 (1-197) x BF863244 (1-701)
Qy 71 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 86
Db 171 CAGCACCATACCTTGCCTCAGTGTCTCACACCAAAATTATCCCATACGGCCACTA 230
Qy 87 ThrGlnMetThrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySer 106
Db 231 ACNAAAGTTACATACACGG-----AAGACACAGCGCGCTTGGCCACCCCTTTGGAGCCG 284
Qy 107 ArgSerLeuThrProCysThrCysGlySerSerAspLeuValThrArgHisAla 126
Db 285 AGAAGCCCGACCGCTGCTCTGGGTCTATCCGATGCTATGCAATCTCCCGTATCAG 344
Qy 127 AspValIle-----ProValArgArgArgGlyAspSerArg----- 138
Db 345 GAGATCATTTTGCATGTGGCTTTAGTCACCCCAAGAGAGCCTGGAGTGGSCATTTATAA 404
Qy 139 -----GlySerLeuLeuSerProArgProIleSer---Tyr 149
Db 405 GAAGGGGACGGAATTCGTTTGGGAAAGTGGAGCGCCCAANGTCTGACCAAGTGCTA 464
Qy 150 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169
Db 465 CTCCAAGCGCAGCAATGGGAGCCTTTCGGGGTGTGGGGTGTCTCTCTTAATGTGTCAG 524
Qy 170 ArgAlaAlaVal-----SerThrArgGlyValAlaLysAla--- 181
Db 525 AAAGAAACCATTTAGTAGTAGAGTGCCTTATCCCGGAGGTGAAGGTGACCTCTAT 584
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArg 195
Db 585 GTGGACCGCATACATCTCGAAGACACACAGGTGCTACGCA 626

RESULT 7
BF304699/c
LOCUS
DEFINITION
601888252Fl NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122276 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 984)
REFERENCE
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI005 row: g column: 13
High quality sequence stop: 646.
Location/Qualifiers
1..984
/organism="Homo sapiens"
/mol_type="mRNA"

```

BASE COUNT  
ORIGIN

[illegible]

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

```

MG1:1477610
seq primer:  -400P from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. 629
/organism="Mus musculus"
/mol_type="mrna"
/db_xref="taxon:10090"
/clone="IMAGE:3977578"
/tissue_type="NK cells (flow-sorted)"
/lab_host="DH10B (Tl-resistant)"
/clone_lib="NCI_CGAP_Sp2"
/notes="Organ: spleen; Vector: pCMV-SPORT6 (Life
Technologies); mRNA made from flow-sorted NK cells, CDNA
made by oligo-dT priming. Directionally cloned. Average
insert size 1.5 kb. Primary library, non-amplified. CDNA
Library Preparation: David B. Krizman, Ph.D."
131 a 156 c 150 g 191 t 1 others
BASE COUNT
ORIGIN

```

Alignment Scores:					
Pred. No.:	17.5	Length:	629		
Score:	95.00	Matches:	46		
Percent Similarity:	39.53%	Conservative:	22		
Best Local Similarity:	26.74%	Mismatches:	58		
Query Match:	9.34%	Indels:	46		
DB:	10	Gaps:	11		
 US-09-965-594-18 (1-197) x BG089727 (1-629)					
QY	45	GluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr-SerIleAsn	64		
		:::      :::      ::   ::			
Db	608	GAAGGACACACAGATCATCCCTGTGCA---AAACAATTTCCTCCAGATATAAAT	552		
QY	65	GlyValLeuTrpThrValTyrHisGly-	73		
		:::			
Db	551	GCT-----ACTATGTTCTCAGGTGACATCATGCTTTAAAGTGCGAGAGTAAGGCC	501		
QY	74	-----AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyr	90		
		:::    :::    :::			
Db	500	AAGAAGTAAGCTGTGAGACCCCTCAAGTTCCCAGATCCAATGCCGGGTGAAGCCA	441		
QY	91	ThrAsnValAspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArgSerLeu---	109		
		::			
Db	440	GNGAATGTG---TCACGTGTGGTGGCTGG------GGGTCAAGGTCCCATCAAT	396		
QY	110	--ThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspVal	128		
		:   :::       ::	:::		
Db	395	GACACTAAGCATCTGCCGCCTGGCGAGAGGTTCAAAGTCTCCAGGAGACGAGGAAA	336		
QY	129	IleProValArgArgArg--	138		
			:::		
Db	335	TGCAAAAAACGTTTCCGATACTACACTGACACCACAGAGATTGTGCTGGAGACTTGAG	276		
QY	139	GlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeu	158		
		:::    :::			
Db	275	---AAAAATAAGACTCCT-	234		
			TTCAAGGTGACTCGGGGGGACCCCTT		
QY	159	LeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAla-----ValSer	174		
		::::	::		
Db	233	GTGTGTGAC----AACCAAGCATATGCACTTTTGCCTATGCAAAAACGGAACAATCTCT	177		
QY	175	ThrArgGlyValAlaLysAlaValAspPheIlePro	186		
		:::			
Db	176	TCAGGATCTTCATAAGTTGTGCACCTCCTCGCG	141		

RESULT 9  
B0542842/c



```

LOCUS      BU542842                905 bp      mRNA      linear      EST 13-SEP-2002
DEFINITION AGENCOURT_10334715 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574789
5' mRNA sequence.
ACCESSION  BU542842
VERSION     BU542842.1 GI:22853325
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 905)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/UTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2770 row: k column: 13
            High quality sequence stop: 633.
            Location/Qualifiers
              1..905
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6574789"
                /tissue_type="carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_40"
            Note: "organ: prostate; Vector: pOTB7; Site_1: XhoI;
            Site_2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
BASE COUNT 201 a 260 c 273 g 171 t
ORIGIN
Alignment Scores:
Pred. No.:      39, 5      Length:      905
Score:          93.50     Matches:     51
Percent Similarity: 31.88%   Conservative: 15
Best Local Similarity: 24.64% Mismatches:    74
Query Match:     9.19%     Indels:      67
DB:              13       Gaps:         9

US-09-965-594-18 (1-197) x BU542842 (1-905)
Qy 28 GluGluGlyCysGlnGlnThrSerGln---ThrGlyArgAspLysAsnGlnValGluGly 46
Db 884 AAGGAAGGGCCCGCAGTCTGTCTCCAGGAAAGGGGACCCGAGACCAAGGAGGAGGC 825
Qy 47 GluValGlnIleValSerThrAlaThrGlnThrPheLeuAla----- 60
Db 824 GGGGCCCTTTCACGGCCCTGTGCACAAAGTGTCCCTTGGGGTGCCCGCCATGTGCCA 765
Qy 61 -----ThrSerIleAsnGlyValLeuTrpThr----- 69
Db 764 CATTCTCGAGCATCGGCAGAACATGTGGTCCGCTTGGCCACAGCAGGACGCC 705
Qy 70 -----ValThrHisGlyAla----- 74
Db 704 AAGTGGGAGGAGGCGATGGTGCACACCTGGGGAGGCGCCCTGGTGCAGAACGACGCCCA 645
Qy 75 -----GlyThrArgThrIle 79

```

```

Db 644 CAGTAGCAGCCCATCCAGGAGAACACCACTCCGGAGGGCCACAGCCCTCTCGACGCCCTG 585
Qy 80 AlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGly 99
Db 584 GCATCCGCCGCCAGCCCTCCATCTCAGCGGGATGTGCACGGGTGAGACAGGAATGCAGGGA 525
Qy 100 TrpGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu 119
Db 524 CGTTCGCCCTAGGTCAGCCCTTCATCCGCCCTGTGTGCTGTGTCAGTGGTCAAGGTG 465
Qy 120 -----TyrLeuValThrArgHisAlaAsp-----ValIleProValArgArg 133
Db 464 CCCTGCCACAGCTGCTGCAAGCCATCCAGGGCTTCGCTGTCTCTCCAGCTCACCTCT 405
Qy 134 Arg-----GlyAspSerArgGlySerLeuLeuSerPro-----Arg 145
Db 404 CGCCTCCAGGGCCAGCCCTTCATCTCTCAGGATCTGGGTAGTCTCTGGGTATCTG 345
Qy 146 ProIleSerTyrLeuLysGlySerGlyProLeuLeuCysProAlaGly----- 163
Db 344 CCTCAGAAAGGGCTGGCAGGCTGTCTGCAGGTGCAGCTGTGCCTCTCTGCTCTCTCTG 285
Qy 164 -----HisAlaValGly 167
Db 284 CGGGTGGCTCAGCGTGCAGGG 264

RESULT 10
BM553374 1733 bp mRNA linear EST 20-FEB-2002
LOCUS     AGENCOURT_6558368 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742981
DEFINITION 5' mRNA sequence.
ACCESSION  BM553374
VERSION     BM553374.1 GI:18792049
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1733)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12761 row: p column: 22
            High quality sequence start: 88
            High quality sequence stop: 539.
            Location/Qualifiers
              1..1733
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5742981"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_119"
            Note: "Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: EcoRV (destroyed); RNA source normal medulla from
            anonymous male age 27. Library is oligo-dT primed and
            directionally cloned (EcoRV site is destroyed upon
            cloning). Average insert size 1.3 kb, insert size range
            0.9-3 kb. Library is normalized and enriched for
            full-length clones and was constructed by C. Gruber
            (Invitrogen). Research Genetics tracking code 013. Note:
            this is a NIH_MGC Library."
BASE COUNT 243 a 673 c 521 g 288 t

```

## ORIGIN

## Alignment Scores:

Pred. No.: 92.2 Length: 1733  
 Score: 93.50 Matches: 56  
 Percent Similarity: 37.26% Conservative: 23  
 Best Local Similarity: 26.42% Mismatches: 78  
 Query Match: 9.13% Indels: 55  
 DB: 12 Gaps: 14

US-09-965-594-18 (1-197) x BM553374 (1-1733)

QY 17 GlyAspThrAlaValAlaGlnGlnThrArg-----GlyGluGlu 29  
 |||||  
 Db 1027 GGGGACTCAGGGGGGCTCTCTCACTCGCGCGGGGTCTCCCAAGGGGGGAGAA 968  
 QY 30 Gly-----CysGlnGluThrSerGlnThrGlyArgAspLysAsnGlnVal 44  
 |||  
 Db 967 GGGTCTTCAGAGGCTTTGTGAGCGCTTTTGGGAGAGCTAGGCCAAAGAAGAGGTC 908  
 QY 45 GluGlyGluValGln-----IleValSerThrAlaThrGlnThrPheLeuAlaThrSer 62  
 |||||  
 Db 907 CCGGGGACCAAAATTCGCTGTTGTCCTCAATTCACCAAGGCT---GTTTCATCCAG 851  
 QY 63 IleAsnGlyValLeu-----TrpThrValThrHisGlyAlaGlyThrArgThr---Ile 79  
 |||||  
 Db 850 GTGGAGGCGCTCTTCGCGGGGTGGCCCTTTGGAGGGGAAGCGCGCGGAGTGGTTT 791  
 QY 80 AlaSerProLysGlyProValThrGlnMetThrThrAsnValAlaPlyAspLeuValGly 99  
 |||||  
 Db 790 CCTCCCGGAGGAGTCCAAAGGGGTTTTC-----CTCCAGGC 746  
 QY 100 -----TrpGlnAlaProGlnGlySerArgSerLeuThrProCys 112  
 |||||  
 Db 745 TCCCGTCCGCGCCCTGCTTCTGGGAAGCCCGCCAGGCGCAGCA----- 701  
 QY 113 ThrCysGlySerSerAspLeuThrValThrArgHisAlaAspValIleProValArg 132  
 |||||  
 Db 700 -----GGCCCGGAGTGGCGAGCTGCTCAGGAGACATGCA----- 668  
 QY 133 ArgArgGlyAspSerArgGlySerLeuLeuSerProArg---ProIleSerTyrLeuLys 151  
 |||||  
 Db 667 CCGCGGGTGATGGCGCCAGCGGTGGCCCATCTTCCCTGGCGCTGGAGAGGCCAGG 608  
 QY 152 GlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAla 171  
 |||||  
 Db 607 GGATGCTCCAGGAGTGCCCGCGGCTGCCAGTGCGTGACGGTGGAAGTGCCCGCGTCT 548  
 QY 172 AlaValSer-----ThrArgGlyValAlaLys-----AlaValAspPheIle 185  
 |||||  
 Db 547 GCTGTGGAACACTGACACATACCACAGGAATGGCCCGGTGCGCCGGAGTCTGCTCTC 488  
 QY 186 ProValGluSerLeuGluThrThrMetArgSerPro 197  
 |||||  
 Db 487 CCA-----CGCAGGGGGGGGGAGTCCG 464

## RESULT 11

## BM402566

## LOCUS

BM402566 528 bp mRNA linear EST 01-JUL-2002  
 SLA005f12\_34513 An expressed sequence tag (EST) collection from the  
 resurrection plant *Selaginella lepidophylla* SLA005f12 5, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

*Selaginella lepidophylla*  
*Selaginella lepidophylla*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;  
*Selaginella*  
 1 (bases 1 to 528)

## REFERENCE

## AUTHORS

## TITLE

Iturriaga, G. and Cushman, J.C.  
 An expressed sequence tag (EST) collection from the resurrection

JOURNAL  
COMMENT

plant *Selaginella lepidophylla*  
 Unpublished  
 Contact: Cushman JC  
 Department of Biochemistry  
 University of Nevada  
 MS200, Reno, NV 89557-0014, USA  
 Tel: 775-784-1918  
 Fax: 775-784-1650  
 Email: jcushman@unr.edu  
 PCR PRIMERS  
 FORWARD: T3 20mer  
 BACKWARD: T7 21mer  
 Plate: 005 row: F column: 12  
 Seq primer: T3 20mer  
 High quality sequence stop: 528.

## FEATURES

## source

1..528  
 /organism="Selaginella lepidophylla"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:59777"  
 /clone="SLA005f12"  
 /tissue\_type="microphyll fronds undergoing desiccation for  
 2.5 h"  
 /dev\_stage="adult"  
 /clone\_lib="An expressed sequence tag (EST) collection  
 from the resurrection plant *Selaginella lepidophylla*"  
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site\_1:  
 EcoRI; Site\_2: XhoI; Library construction was performed  
 according to manufacture's (Stratagene, Inc.) recommended  
 protocol for the Lambda UniZapXR vector and cDNA synthesis  
 kit."  
 129 a 125 c 137 g 137 t

BASE COUNT  
ORIGIN

## Alignment Scores:

Pred. No.: 21.9 Length: 528  
 Score: 93.00 Matches: 37  
 Percent Similarity: 42.98% Conservative: 15  
 Best Local Similarity: 30.58% Mismatches: 43  
 Query Match: 9.14% Indels: 26  
 DB: -12 Gaps: 4

US-09-965-594-18 (1-197) x BM402566 (1-528)

QY 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113  
 |||||  
 Db 53 GACAAGGATGTAGCGGTGCTGAAGATCGATGCTCAAGCAACAGATCTCAGCGCAATACCC 112  
 QY 114 CysGlySerSerAspLeuTyrLeuVal----- 122  
 |||||  
 Db 113 CTTGGAAGTCTCGCGATCTGCTTGTGCCAGAGGTGTATGCTATCGGTAATCCTTTT 172  
 QY 123 -----ThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138  
 |||||  
 Db 173 GGATTTGGATCATACGCTGACACAGCGCTCATCTTCCGAAGGGAGATTACT--- 229  
 QY 139 GlySerLeuLeuSerProArgProIleSerTyrLeu----- 150  
 |||||  
 Db 230 ---TCAGCGCTTAATGGTGTCCATCCNAGACGTGTATCCAGACAGATGCGCGTATTAT 286  
 QY 151 LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170  
 |||||  
 Db 287 CCGTGGACACAGCGGGGTCGCTATTGGACAGTCTCTGGAAATTTGATAGGCATCAACACT 346  
 QY 171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189  
 |||||  
 Db 347 GCTATATATTTCTCCGCTCTGGCGCTCATCAGCGGTTCATTCATTCAGTTGACACG 406  
 QY 190 Leu 190  
 |||||  
 Db 407 GTT 409

## RESULT 12

**BX238988/c**  
**LOCUS**  
**DEFINITION** Danio rerio genomic clone DREY-283L13, linear GSS 29-JAN-2003  
**ACCESSION** BX238988  
**VERSION** BX238988.1 GI:28161322  
**KEYWORDS** GSS.  
**SOURCE** Danio rerio (zebrafish)  
**ORGANISM** Danio rerio  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
**AUTHORS** 1 (bases 1 to 644)  
**TITLE** Humphray,S.J., Huckle,E. and Durham,J.L.  
**JOURNAL** Direct Submission  
**COMMENT** Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 This sequence was generated from the T7 end of BAC 283L13. 283L13 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: [http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/).  
**FEATURES** Location/Qualifiers  
 source  
 1..644  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DREY-283L13"  
 /tissue\_type="Testis"  
 /note="vector pIndigoBAC-536"  
**BASE COUNT** 129 a 212 c 176 g 127 t  
**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 28.4 Length: 644  
 Score: 93.00 Matches: 37  
 Percent Similarity: 44.72% Conservative: 18  
 Best Local Similarity: 30.08% Mismatches: 52  
 Query Match: 9.14% Indels: 16  
 DB: 29 Gaps: 7  
 US-09-965-594-18 (1-197) x BX238988 (1-644)  
 QY 68 TrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThr 87  
 DB 405 TGGACAGCATCTCGCAGAGTGGAGCTCGAGGGCGGTGTCGCCGTCACCATGAGC 346  
 QY 88 GlnMetTyr-----ThrAsnValAspLysAspLeuValGlyTrpGlnAla 102  
 DB 345 AGCAGCTGGTCATCGTGGTGTCCAGCAGTCTTGGAAAGGAC-----TGGAGATCC 295  
 QY 103 ---ProGlnGlySerArg-----SerLeuThrProCysThrCysGlySerSerAsp 118  
 DB 294 AGACCGTTGGGGCAAAAGCGGAGGAGTGTGTCACCATGCGCTTTAAAAAAGGAGAAA 235  
 QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138  
 DB 234 TTATTGTGATTACCATGGGAGGAGTGTGCGCAGGAGGCGCTTAGGAGAGCGGAGG 175  
 QY 139 GlySerLeuLeuSerProArgPro-----IleSerTyrLeuLysGlySerSerGlyGly 156  
 DB 174 CCTCTCCCGGTGCTCCTCTCTCATATGTTCT---TTAAAGGCGCTTGGTGGGAGGA 118  
 QY 157 ProLeuLeu---CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThr 175  
 DB 117 CCTTTGCTGATGCCCGCATGCTTGGCCCTGCTGACCCGGGCATGGAACCTTCG 58  
 QY 176 ArgGlyVal 178  
 DB 57 GAAGCGCTTA 49  
**RESULT 13**  
**BZ342381/c**  
**LOCUS**

**DEFINITION** ic83b11.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum bicolor genomic clone ic83b11 5', genomic survey sequence.  
**ACCESSION** BZ342381  
**VERSION** BZ342381.1 GI:24742983  
**KEYWORDS** GSS.  
**SOURCE** Sorghum bicolor (sorghum)  
**ORGANISM** Sorghum bicolor  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
**AUTHORS** 1 (bases 1 to 701)  
**TITLE** Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.  
**JOURNAL** Genomic shotgun sequences from Sorghum bicolor (methyl-filtered) Unpublished  
**COMMENT** Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: ic83 row: b column: 11  
 Seq primer: -21M13UnivFwd  
 Class: shotgun  
 High quality sequence stop: 701.  
**FEATURES** Location/Qualifiers  
 source  
 1..701  
 /organism="Sorghum bicolor"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4558"  
 /clone="ic83b11"  
 /lab\_host="JM107 or DH5a"  
 /clone\_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"  
 /note="Site 1: Xba I; Site 2: Xba I; The vector was digested with Xba I and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed in either JM107 or DH5a."  
**BASE COUNT** 108 a 251 c 232 g 110 t  
**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 31.7 Length: 701  
 Score: 93.00 Matches: 48  
 Percent Similarity: 36.69% Conservative: 14  
 Best Local Similarity: 28.40% Mismatches: 56  
 Query Match: 9.14% Indels: 51  
 DB: 29 Gaps: 8  
 US-09-965-594-18 (1-197) x BZ342381 (1-701)  
 QY 60 AlaThrSerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThr--- 78  
 DB 500 GCCCATGCGGGGGTGTGTTTTTGAACGTACCGCGCGGGGAGGACGACAGGC 441  
 QY 79 -----IleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAsp 96  
 DB 440 GTGGCGGTGGCCCGCCCGCGCGCGCGTGGGAATTATGCCGCATTAACCTCGGCAC 381  
 QY 97 LeuValGlyTrpGlnAlaProGlnGlySerArg----SerLeuThrProCysThr----- 113  
 DB 380 CTACGCGGAGCGGAGACAGAGGTGCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTC 321  
 QY 114 -----CysGlySerSerAsp 118  
 DB 320 CACCCAGTCCCGCGCGCGCGCGCGGTACGGGACAGGAACTGGTGGTGAAGATCAC 261  
 QY 119 LeuTyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAsp----- 136



## ORIGIN

```
Alignment Scores: 40.3 Length: 772
Pred. No.: 92.50 Matches: 47
Score: 40.57% Conservative: 24
Percent Similarity: 26.86% Mismatches: 59
Best Local Similarity: 9.10% Indels: 46
Query Match: 29 Gaps: 10
DB:

US-09-965-594-18 (1-197) x CC406704 (1-772)

QY 39 ArgAspLysAsnGlnVal---GluGlyGluValGlnIleValSerThrAlaThrGlnThr 57
Db 633 AGATGGCGAAGCAGCTAAACACAGGGCTATGTACAAATTGTGACG-----AGCCTATG 580
QY 58 PheLeuAlaThrSerIleAsnGlyVal---LeuTrpThrValTyrHisGlyAlaGlyThr 76
Db 579 TACGAGGCCACCGCTGTCCCGCTATTCTATCTGGAGG-----TCA 538
QY 77 ArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAsp 96
Db 537 CAGACTTTCTGACGATGATCAAGGTGACACAGATGATGACCAGG-----AAGAGC 484
QY 97 LeuValGlyTyrP---GlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 483 AGGACCCAGTGGTTTCCCTTACATGCAATAACTGGGATTAGAGGGAGGACACCATGCGAGC 424
QY 114 CysGlySerSerAspLeu-----TyrLeuValThrArgHisAlaAsp 127
Db 423 TCGGGGTAGTCTCTCAACGGTCAAGGAGCTGCTGGCCCTACTTGACACGGGTTTCAACACATA 364
QY 128 ValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIle 147
Db 363 ACTTCATCAACTGCACAGCGCGCA-CAACAGCGTGGGTTACTTTGGACCCACACACAGGT 305
QY 148 SerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGly 167
Db 304 CGCCATGTCAAGGTGGCAAAATGGAGACCCAGTTTCTGCCAG-----GGA 260
QY 168 IlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe----- 184
Db 259 GTAACTCGTCCGCGCA-----GCCATTGACATCAACAAGGAG 224
QY 185 -----IleProValGluSerLeuGlu 191
Db 223 AAGTTCACCATTGAGGCATATGCAATTCCTTGGATACATTGAG 179
```

Search completed: August 31, 2003, 04:27:41  
Job time : 1917.31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 HMKKGSVIVGRINLSGDTA.....YAKAYDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 15872573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	100.0	197	AA15224	Hepatitis C virus
2	1010	99.0	197	AA15225	Hepatitis C virus
3	1005	98.5	197	AA15223	Hepatitis C virus
4	990	97.1	197	AA15222	Hepatitis C virus
5	973	95.4	197	AA15221	Hepatitis C virus
6	946	92.7	197	AA15226	Hepatitis C virus
7	939	92.1	195	AA15220	Hepatitis C virus
8	912	89.4	195	AA15212	Hepatitis C virus
9	885.5	86.8	665	AA1524943	HCV NS4A-NS3 compl

10	882.5	86.5	665	20	AA1524947	HCV NS4A-NS3 compl
11	881.5	86.4	665	20	AA1524942	HCV NS4A-NS3 compl
12	878.5	86.1	216	20	AA1517880	HCV NS4A-NS3 compl
13	878.5	86.1	665	20	AA1524946	HCV NS4A-NS3 compl
14	877.5	86.0	665	20	AA1524941	HCV NS4A-NS3 compl
15	875.5	85.8	216	20	AA1517884	HCV NS4A-NS3 compl
16	874.5	85.7	216	20	AA1517879	HCV NS4A-NS3 compl
17	874.5	85.7	665	20	AA1524945	HCV NS4A-NS3 compl
18	873.5	85.6	665	20	AA1524940	HCV NS4A-NS3 compl
19	873.5	85.6	671	20	AA1524948	HCV NS4A-NS3 compl
20	871.5	85.4	216	20	AA1517883	HCV NS4A-NS3 compl
21	870.5	85.3	216	20	AA1517878	HCV NS4A-NS3 compl
22	870.5	85.3	665	20	AA1524944	HCV NS4A-NS3 compl
23	870.5	85.3	671	20	AA1524949	HCV NS4A-NS3 compl
24	870	85.3	215	20	AA1517890	HCV NS4A-NS3 compl
25	867.5	85.0	216	20	AA1517882	HCV NS4A-NS3 compl
26	867.5	85.0	216	20	AA1517885	HCV NS4A-NS3 compl
27	866.5	85.0	216	20	AA1517877	HCV NS4A-NS3 compl
28	864	84.7	215	20	AA1517887	HCV NS4A-NS3 compl
29	863.5	84.7	215	20	AA1517881	HCV NS4A-NS3 compl
30	863.5	84.7	216	20	AA1517885	HCV NS4A-NS3 compl
31	859	84.2	213	20	AA1517888	HCV NS4A-NS3 compl
32	859	84.2	631	20	AA153482	HCV NS3 protein.
33	858.5	84.2	191	21	AA1544728	Hepatitis C virus
34	858.5	84.2	3011	19	AA1577397	Hepatitis C virus
35	858.5	84.2	3011	24	ABP71460	Amino acid sequenc
36	858.5	84.2	3012	23	AAU99289	Hepatitis C virus
37	855.5	83.9	3011	14	AA140120	HCV genomic amino
38	854.5	83.8	687	16	AA1579223	pHCV150-encoded se
39	854.5	83.8	1648	16	AA1579221	pHCV176-encoded se
40	854.5	83.8	1766	10	AA152041	Sequence encoded i
41	854.5	83.8	1786	10	AA1520158	Protein sequence o
42	854.5	83.8	2261	10	AA1520164	Peptide encoded by
43	854.5	83.8	2301	10	AA1520167	Sequence encoded i
44	854.5	83.8	2436	10	AA152050	Sequence encoded i
45	854.5	83.8	2436	10	AA1520288	Peptide encoded by

#### ALIGNMENTS

RESULT 1  
AA15224  
ID AAB15224 standard; protein; 197 AA.  
XX  
AC AAB15224;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fuslon protease #6.  
XX  
HE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
WPI; 2000-465976/40.  
DR N-PSDB; AAA73333.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 PS  
 XX Claim 23; Fig 16; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1020; DB 21; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-98;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60  
 QY 61 TSINGVLTWVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGWQAPQGSRSLSLPTCTCGSSDLY 120  
 DB 61 TSINGVLTWVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGWQAPQGSRSLSLPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2  
 AAB15225  
 ID AAB15225 standard; protein; 197 AA.  
 AC AAB15225;  
 DT 19-DEC-2000 (first entry)  
 XX Hepatitis C virus NS4A-NS3 fusion protease #7.  
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; muten.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX WO200040707-A1.  
 XX 13-JUL-2000.  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 PI WPI: 2000-465976/40.  
 DR N-PSDB; AAA73334.  
 DR

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 PS  
 XX Claim 23; Fig 17; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 99.0%; Score 1010; DB 21; Length 197;  
 Best Local Similarity 99.5%; Pred. No. 3.6e-97;  
 Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAQOTRGEGCGCKTSHTGRKNOVEGEVQIVSTATOTFLA 60  
 QY 61 TSINGVLTWVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGWQAPQGSRSLSLPTCTCGSSDLY 120  
 DB 61 TSINGVLTWVYHGAGTRTIAAPKGPVTOMYTNVDKDLVGWQAPQGSRSLSLPTCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLGSSGGPLLCPAGHAVGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3  
 AAB15223  
 ID AAB15223 standard; protein; 197 AA.  
 AC AAB15223;  
 DT 19-DEC-2000 (first entry)  
 XX Hepatitis C virus NS4A-NS3 fusion protease #5.  
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; muten.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX WO200040707-A1.  
 XX 13-JUL-2000.  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 PI WPI: 2000-465976/40.  
 DR N-PSDB; AAA73332.  
 DR

XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
PS  
XX Claim 23; Fig 15; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
XX variant.  
XX  
SQ Sequence 197 AA;  
Query Match 98.5%; Score 1005; DB 21; Length 197;  
Best Local Similarity 98.5%; Pred. No. 1.2e-96;  
Matches 194; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
DB 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
QY 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTCGSSDLY 120  
DB 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAVSTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAVSTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 4  
AAB15222  
ID AAB15222 standard; protein; 197 AA.  
XX  
AC AAB15222;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #4.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO2000040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI; 2000-465976/40.

DR N-PSDB; AAA73331.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
PS  
XX Claim 23; Fig 14; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
XX variant.  
XX  
SQ Sequence 197 AA;  
Query Match 97.1%; Score 980; DB 21; Length 197;  
Best Local Similarity 97.0%; Pred. No. 4.4e-95;  
Matches 191; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
DB 1 MKKGSVVIVGRINLSGDTAYAAQTGEGCGCKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
QY 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTCGSSDLY 120  
DB 61 TSINGVLTWVYHGAGTRTITASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAVSTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGLSPRPISYLGKSSGGPLLCPAGHAGVIFRAAVSTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 5  
AAB15221  
ID AAB15221 standard; protein; 197 AA.  
XX  
AC AAB15221;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #3.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO2000040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI; 2000-465976/40.



DR WPI: 2000-465976/40.  
 DR N-PSDB: AAA73330.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 PS Claim 23; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 95.4%; Score 973; DB 21; Length 197;  
 Best Local Similarity 95.4%; Pred. No. 2.6e-93; Indels 0; Gaps 0;  
 Matches 188; Conservative 2; Mismatches 7;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLVTYHAGATRTIASPKGPVQMTNVDKDLVGHQAPGQSRSLTPCTCGSSDLY 120  
 DB 61 TCINGVCVTVYHAGATRTIASPKGPVQMTNVDKDLVGHQAPGQSRSLTPCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLAGSSGGLPCPAGHAVGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLAGSSGGLPCPAGHAVGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 6  
 AAB15226  
 ID AAB15226 standard; protein; 197 AA.  
 AC AAB15226;  
 DT 19-DEC-2000 (first entry)  
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.  
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.  
 DR N-PSDB: AAA73335.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 PS Example 5; Fig 18; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0  
 CC wild-type sequence.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 92.7%; Score 946; DB 21; Length 197;  
 Best Local Similarity 94.4%; Pred. No. 1.7e-90;  
 Matches 186; Conservative 0; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTRGEGCQKTSHTGRDKNOVEGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLVTYHAGATRTIASPKGPVQMTNVDKDLVGHQAPGQSRSLTPCTCGSSDLY 120  
 DB 61 TCINGVCVTVYHAGATRTIASPKGPVQMTNVDKDLVGHQAPGQSRSLTPCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLAGSSGGLPCPAGHAVGIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGSLLSPRPISYLAGSSGGLPCPAGHAVGIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 7  
 AAB15220  
 ID AAB15220 standard; protein; 195 AA.  
 AC AAB15220;  
 DT 19-DEC-2000 (first entry)  
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
 XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB; AAY73329.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -

XX Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.

XX Sequence 195 AA:

Query Match 92.1%; Score 939; DR 21; Length 195;  
 Best Local Similarity 93.4%; Pred. No. 9.2e-90;  
 Matches 184; Conservative 3; Mismatches 8; Indels 2; Gaps 1;  
 QY 1 MKKGGSVVIGRINLSGDTAYAQOTRGEQGCOKTSHTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGGSVVIGRIVLNG--AYAQOTRGEQGCOKTSHTGRDKNQVEGEVQIVSTATQTFLA 58  
 QY 61 TSINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 120  
 DB 59 TCINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 118  
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFPRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFPRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 8  
 AAB15212  
 ID AAB15212 standard; protein: 195 AA.

XX AAB15212;

DT 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease #1.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer.

XX Hepatitis C virus.

OS Synthetic.

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB; AAY73328.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -

XX Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C  
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
 CC essential for the replication of the virus, acting to cleave its  
 CC replicative proteins from the polyprotein produced from the HCV genome.  
 CC Inhibitors of the two proteins should be effective as antiviral  
 CC treatments of HCV infection. This is useful as HCV can lead to chronic  
 CC liver disease such as cirrhosis, liver failure and liver cancer. The  
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
 CC proteins which can be used to identify inhibitors of this type, as well  
 CC as enabling structural studies of the protease and protease:inhibitor  
 CC complexes.

XX Sequence 195 AA:

Query Match 89.4%; Score 912; DB 21; Length 195;  
 Best Local Similarity 92.4%; Pred. No. 6.1e-87;  
 Matches 182; Conservative 1; Mismatches 12; Indels 2; Gaps 1;  
 QY 1 MKKGGSVVIGRINLSGDTAYAQOTRGEQGCOKTSHTGRDKNQVEGEVQIVSTATQTFLA 60  
 DB 1 MKKGGSVVIGRIVLNG--AYAQOTRGLLCITSLTGRDKNQVEGEVQIVSTATQTFLA 58  
 QY 61 TSINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 120  
 DB 59 TCINGVLTWVYHGAGTTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLY 118  
 QY 121 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFPRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFPRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 9  
 AAY24943  
 ID AAY24943 standard; protein: 665 AA.

XX AAY24943;

DT 07-SEP-1999 (first entry)

DE HCV NS4A-NS3 complex SEQ ID NO:14.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.

XX Hepatitis C virus.

OS Synthetic.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

```

XX PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 90-92; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;

Query Match 86.8%; Score 885.5; DB 20; Length 665;
Best Local Similarity 85.7%; Pred. No. 2e-83;
Matches 168; Conservative 15; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTGRDKNQVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSHTGRDKNQVEGEVQIVSTATOTFLAT 81
QY 62 SINGVLMTVTHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGGSRLTPCTCGSSDLYL 121
DB 82 CVNGVCVTVTHGAGSKTLAGPKGPIQMTYTNVDQDLVGWQAPPGARSRLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRKDSRGSLLSPRPISYLKSGSGGPLLCPCAGHAGVGFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRKDSRGSLLSPRPVSYLKGSGGPLLCPSGHAGVGFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 10
AAY24947
ID AAY24947 standard; Protein; 665 AA.
XX AC AAY24947;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:18.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```

```

PA (SCHE ) SCHERING CORP.
XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX DR WPI; 1999-385385/32.
XX PT New hepatitis C virus covalent complexes
XX PS Claim 6; Page 100-102; 21pp; English.
XX CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;

Query Match 86.5%; Score 882.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 4.1e-83;
Matches 167; Conservative 16; Mismatches 10; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTGRDKNQVEGEVQIVSTATOTFLAT 61
DB 22 GSVVIVGRILSGSGSITAYSQOTRGLGCKKTSHTGRDKNQVEGEVQIVSTATOTFLAT 81
QY 62 SINGVLMTVTHGAGTRTIA SPKGPVTOMYTNVDKDLVGWQAPGGSRLTPCTCGSSDLYL 121
DB 82 CVNGVCVTVTHGAGSKTLAGPKGPIQMTYTNVDQDLVGWQAPPGARSRLTPCTCGSSDLYL 141
QY 122 VTRHADVIPVRRKDSRGSLLSPRPISYLKSGSGGPLLCPCAGHAGVGFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRKDSRGSLLSPRPVSYLKGSGGPLLCPSGHAGVGFRAAVCTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217

RESULT 11
AAY24942
ID AAY24942 standard; Protein; 665 AA.
XX AC AAY24942;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:13.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```



XX WPI: 1999-385385/32.  
 DR New hepatitis C virus covalent complexes  
 XX  
 PT Claim 6; Page 97-99; 21lpp; English.  
 XX  
 PS The present invention describes a covalent hepatitis C virus (HCV)  
 XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.  
 XX Sequence 665 AA;  
 SQ

Query Match 86.1%; Score 878.5; DB 20; Length 665;  
 Best Local Similarity 85.2%; Pred. No. 1.1e-82;  
 Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;  
 QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTGRDNQVGEVQIVSTATQFLAT 61  
 DB 22 GSVVIVGRIRIILSGSGSITAYSQOTRGLGCKITSLTGRDNQVGEVQIVSTATQFLAT 81  
 QY 62 SINGVLWTVYHGAGTRTITASPKGPVTQMTYNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121  
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYNVDODLVGWQAPPGARSLTPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRGRDSRGLSPRPISYLGSGGGLPCPAGHANGVIFRAAVSTRGVAKA 181  
 DB 142 VTRHADVIPVRRGRDSRGLSPRPISYLGSGGGLPCPAGHANGVIFRAAVSTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETMRSP 217

RESULT 14  
 AAY24941  
 ID AAY24941 standard; Protein: 665 AA.  
 XX  
 AC AAY24941;  
 DT 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex SEQ ID NO:12.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WPI: 1999-385385/32.

DR WPI: 1999-385385/32.  
 XX New hepatitis C virus covalent complexes  
 XX  
 PT Claim 6; Page 85-87; 21lpp; English.  
 XX  
 PS The present invention describes a covalent hepatitis C virus (HCV)  
 XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence represents a specifically claimed example of the above  
 CC complex. The covalent NS4A-NS3 complexes are useful for structural  
 CC determination and determination of mode of binding of HCV inhibitors by  
 CC NMR spectroscopy. They can also be used for detecting inhibitors of the  
 CC protease activity, the helicase activity and the ATPase activity of NS3.  
 CC The covalent NS4A-NS3 complexes are more soluble, stable and active than  
 CC the non-covalent protease-peptide complexes previously available.  
 XX Sequence 665 AA;  
 SQ

Query Match 86.0%; Score 877.5; DB 20; Length 665;  
 Best Local Similarity 85.2%; Pred. No. 1.4e-82;  
 Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;  
 QY 5 GSVVIVGRINLSGD---TAYAQOTRGEGCGCKTSHTGRDNQVGEVQIVSTATQFLAT 61  
 DB 22 GSVVIVGRIRIILSGSGSITAYSQOTRGLGCKITSLTGRDNQVGEVQIVSTATQFLAT 81  
 QY 62 SINGVLWTVYHGAGTRTITASPKGPVTQMTYNVDKDLVGWQAPGQSRSLTPTCTCGSSDLYL 121  
 DB 82 CVNGVCWTVYHGAGSKTLAGPKGPITQMTYNVDODLVGWQAPPGARSLTPTCTCGSSDLYL 141  
 QY 122 VTRHADVIPVRRGRDSRGLSPRPISYLGSGGGLPCPAGHANGVIFRAAVSTRGVAKA 181  
 DB 142 VTRHADVIPVRRGRDSRGLSPRPISYLGSGGGLPCPAGHANGVIFRAAVSTRGVAKA 201  
 QY 182 VDFIPVESLETTMRSP 197  
 DB 202 VDFVPVESMETMRSP 217

RESULT 15  
 AAY17884  
 ID AAY17884 standard; Protein: 216 AA.  
 XX  
 AC AAY17884;  
 DT 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex SEQ ID NO:8.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WPI: 1999-385385/32.



Result No.	Score	Query Match	Length	DB	ID	Description
1	854.5	83.8	3011	1	GNVC3	genome polyprotein
2	853.5	83.7	3011	1	S40770	genome polyprotein
3	848.5	83.2	3011	1	GNVCV	genome polyprotein
4	836.5	82.0	3010	1	GNVTV	genome polyprotein
5	827.5	81.1	3010	1	A45573	genome polyprotein
6	823.5	80.7	3010	1	GNVTC	genome polyprotein
7	823.5	80.7	3010	1	GNVCV3	genome polyprotein
8	811.5	79.6	3010	1	JS8030	genome polyprotein
9	743.5	72.9	3014	1	JC5620	genome polyprotein
10	675	66.2	3033	1	GNVJ8	genome polyprotein
11	673	66.0	3033	1	QJ1303	genome polyprotein
12	251	24.6	3005	2	T08841	polyprotein - dour
13	245	24.0	2970	2	T08839	polyprotein - marm
14	85.5	8.4	209	2	H31144	probable aromatic
15	85.5	8.4	398	2	R71284	probable periplasm
16	84.5	8.3	716	2	G83612	hypothetical prote
17	83.5	8.2	590	2	B81104	nitrate/nitrite se
18	83.5	8.2	590	2	C81911	nitrate/nitrite se
19	83	8.1	377	2	A75335	hypothetical prote
20	83	8.1	452	2	I39383	angio-associated m
21	83	8.1	1615	2	JE0372	low density lipopr
22	80.5	7.9	479	2	H70847	probable oxidoredu
23	79.5	7.8	433	2	H97199	htra-like serine p
24	79.5	7.8	1049	2	T42045	beta transducin-li
25	78.5	7.7	394	2	F95973	probable sugar upt
26	78.5	7.7	485	2	B71360	hypothetical prote
27	78.5	7.7	764	2	A9448	irregular chiasm C
28	78.5	7.7	846	2	T04533	hypothetical prote
29	78	7.6	322	2	D87603	glycosyl transfera





A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A:Reference number: A40244; MUID:92220206; PMID:1314449  
 A:Accession: A40244  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <CHE>  
 A:Cross-references: GB:M4754  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 F:115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEE>  
 F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Product: hepatitis C virus NS2 #status predicted <NS2>  
 F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F:1316-1319/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 82.08; Score 836.5; DB 1; Length 3010;  
 Best Local Similarity 78.4%; Pred. No. 4.4e-68;  
 Matches 160; Conservative 18; Mismatches 17; Indels 9; Gaps 1;

Qy 3 KKGWVIVGRIN-----LSGDTAYAOOTRGECCQKTSHTGRDKNQVEGEVIVST 53  
 Db 1005 RRGREILLGPADSLGEGWALLAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVST 1064

Qy 54 ATQTFLATSLNGVLWTVYHGAGTRTIASPKGPVTQMTYNDKLVGQWAPQGSRSITPCT 113  
 Db 1065 ATQSFATCNGVCTVYHGAGSKTLAGPKGPITQMTYNDQDLVGMHAPPGARSITPCT 1124

Qy 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 173  
 Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 1184

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5  
 A45573  
 genome polyprotein - hepatitis C virus (strain JT)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C:Accession: A45573  
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Virus Res. 23, 39-53, 1992  
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
 A:Reference number: A45573; MUID:92295714; PMID:1318627  
 A:Accession: A45573  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3010 <TAN>  
 A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613  
 A:Experimental source: HCV-JT  
 A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEE>  
 F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 81.18; Score 827.5; DB 1; Length 3010;  
 Best Local Similarity 77.0%; Pred. No. 2.9e-67;  
 Matches 157; Conservative 20; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KKGWVIVGRIN-----LSGDTAYAOOTRGECCQKTSHTGRDKNQVEGEVIVST 53  
 Db 1005 RRGREILLGPADSLGEGWALLAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVST 1064

Qy 54 ATQTFLATSLNGVLWTVYHGAGTRTIASPKGPVTQMTYNDKLVGQWAPQGSRSITPCT 113  
 Db 1065 ATQSFATCNGVCTVYHGAGSKTLAGPKGPITQMTYNDQDLVGMHAPPGARSITPCT 1124

Qy 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 173  
 Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 1184

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 6  
 GNMVTC  
 genome polyprotein - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
 C:Accession: A38465  
 R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, J. Virol. 65, 1105-1113, 1991  
 A:Title: Structure and organization of the hepatitis C virus genome isolated from hu  
 A:Reference number: A38465; MUID:91140698; PMID:1847440  
 A:Accession: A38465  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <TAN>  
 A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEE>  
 F:730-1006/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 80.7%; Score 823.5; DB 1; Length 3010;  
 Best Local Similarity 76.5%; Pred. No. 6.8e-67;  
 Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KKGWVIVGRIN-----LSGDTAYAOOTRGECCQKTSHTGRDKNQVEGEVIVST 53  
 Db 1005 RRGREILLGPADSLGEGWALLAPITAYAOOTRGLGCIITSLTGRDKNQVEGEVIVST 1064

Qy 54 ATQTFLATSLNGVLWTVYHGAGTRTIASPKGPVTQMTYNDKLVGQWAPQGSRSITPCT 113  
 Db 1065 ATQSFATCNGVCTVYHGAGSKTLAGPKGPITQMTYNDQDLVGMHAPPGARSITPCT 1124

Qy 114 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 173  
 Db 1125 CGSSDLVLTTRHADVIPVRRGDSRGLSPRISYLYKSSGGPLLCPCGHAVGIFRAAV 1184

A: Variety: isolate JK1

C: Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001

C: Accession: S18030; S33570; A48332; S18029

C: Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991

A: Description: A whole genome of hepatitis C virus cDNA was isolated from a single p.p.  
A: Reference number: S18028

A: Accession: S18030

A: Molecule type: genomic RNA

A: Residues: 1-3010 <R0N>

A: Cross-references: EMBL: X61596; NID: g59478; PIDN: CAA43793.1; PID: g59479

A: Experimental source: Isolate JK1 from an individual

R: Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993

A: Title: Sequence analysis of putative structural regions of hepatitis C virus isolate JK1

A: Reference number: A48332; MUID: 93119270; PMID: 8380322

A: Accession: S33570

A: Molecule type: genomic RNA

A: Residues: 1-547, 'T', '549-621', 'V', '623-624', 'S', '626-652', 'DL', '655-761', 'T', '763-782' <H0W>

A: Cross-references: EMBL: X61591

A: Note: this sequence is inconsistent with the nucleotide translation

A: Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 45 as Trp, and TTC for residue 771 as Ser

A: Note: sequence extracted from NCBI backbone (NCBI:121747, NCBIP:121748)

C: Superfamily: hepatitis C virus genome polyprotein

C: Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se

F: 2-115/Product: capsid protein C #status predicted <CPC>

F: 116-191/Product: envelope protein M #status predicted <EPM>

F: 192-389/Product: major envelope protein E #status predicted <MEE>

F: 390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F: 730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F: 1007-1615/Product: hepacivirin #status predicted <NS3>

F: 1230-1237/Region: nucleotide-binding motif A (P-loop)

F: 1312-1317/Region: nucleotide-binding motif B

F: 1316-1319/Region: DEXH motif

F: 1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F: 1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F: 2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F: 196, 209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate

Query Match 79.6%; Score 811.5; DB 1; Length 3010;  
Best Local Similarity 76.0%; Pred. No. 8.7e-66;  
Matches 155; Conservative 20; Mismatches 20; Indels 9; Gaps 1;

Qy 3 KKGSVIVGRIN-----LSGDTAYAQOTREGQCQKTSHTGRDNQVEGEVQIVST 53  
Db 1005 RRREILLFGADFGREGQWLLAPITATISYQOTRGLFGCIVTSLTGRDNQVEGEAQVYST 1064

Qy 54 ATQTFATSLNGVLWTVYHGAGTRTASPKGPVQMTYNDKDLVQWQAPGSSSLPCT 113  
Db 1065 ATQSFATCVNGCVTWYHNGSKYTLACPKGPINQMTYNDQDLVQWQAPSGAASLPCT 1124

Qy 114 CGSDDLVLVTRHADVIPRRRGDSRGSLLSPRPISYLGKSGGGPLLCAGHAGVIFRAAV 173  
Db 1125 YGSSDDLVLVTRHADVIPRRRGDSRGSLLSPRPISYLGKSGGGPLLCPSGHAGVIFRAAV 1184

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETMRSP 1208

RESULT 9  
JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N: Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C: Species: hepatitis C virus

C: Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

A: Reference number: JC5620; MUID: 97366593; PMID: 9223423

A: Accession: JC5620

R: Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A: Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A:Accession: PQ0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <XAT>  
C:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Atp; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein E; polyprotein; serine proteinase; tra  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis virus #status predicted <NS3>  
F:1234-1241/Region: nucleotide-binding motif A (P-loop)  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038

Query Match 66.2% Score 675; DB 1: Length 3033;  
Best Local Similarity 69.8%; Pred. No. 3e-53;  
Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;

QY 19 TATAOQTRGEGCQKTSHTGRDNQVGEVQIVSTATQTFLATISINGVLTVTHGACTRT 78  
Db 1034 TATQOQTRGLGAIIVSLTGRDNQVGEVQIVSTATQTFLATISINGVLTVTHGAGNKT 1093  
QY 79 IASPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCTGSSDLXLVTRHADVIPYRRRGDSR 138  
Db 1094 LAGPKGPVTOMYTNVDKDLVGWQAPQGSRLTPTCTGSSDLXLVTRHADVIPYRRRGDSR 1153  
QY 139 GSLLSPRPISYLVKSGGGLPLCPAGHANGIFRAAVSTRGVAKAVDFIPVSELTMRSP 197  
Db 1154 GALLSPRLSTLKGSGGPPVLCRSHGAVGIFRAAVSTRGVAKAVDFIPVSELTMRSP 1212

RESULT 11

genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C:Accession: JQ1303  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurei, K.; Iizuka, H.; Machida, A.; Miyakawa, J. Gen. Virol. 72, 2697-2704, 1991  
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a  
A:Reference number: JQ1303; MUID:9204440; PMID:1658196  
A:Accession: JQ1303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651  
A:Experimental source: isolate HC-J6 from a Japanese individual  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Atp; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038

Query Match 66.0% Score 673; DB 1: Length 3033;  
Best Local Similarity 68.7%; Pred. No. 4.6e-53;  
Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;

QY 19 TAYAOOTRGEQCKQTSHTGRDKNOVEGEVOIVSTATOTFLAITSINGVLTWYVHGAGT 78  
 Db 1034 TAYAOOTRGLLCTIVVSMGTGRDKTRQAGEIQVLTSTVTSFLCTISGVLTVYHGACNKT 1093

QY 79 IASPKGPVTOMYTNVDKLVGNOAPOGSGSLTPTCGSSDLYLVTRHADVIVPVRGGDSR 138  
 Db 1094 LAGSRGPVTQMYSSAEGDLVGPSPGTSKLEPCTCGGAVDLYLVTRNADVIVPARRGDKR 1153

QY 139 GSLSPRPTSYLKGSGGGLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRSP 197  
 Db 1154 GALLSPRLSTLKGSGGGLVPCRGHAGVIFRAAVSTRGVAKSDFIPVELDIVTNSP 1212

RESULT 12  
 T08841  
 C:Species: douroucouli hepatitis GB virus A  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C:Accession: T08841  
 R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A:Reference number: Z16486; MUID:98120818; PMID:9460920  
 A:Accession: T08841  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3005 <ERK>  
 A:Cross-references: EMBL:AF023425; NID:q2828599; PIDN:AAC40502.1; PID:q2828600  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: polyprotein

Query Match 24.6%; Score 251; DB 2; Length 3005;  
 Best Local Similarity 34.1%; Pred. No. 2,6e-14;  
 Matches 56; Conservative 29; Mismatches 69; Indels 10; Gaps 3;

QY 33 KTSHTGRDKNOVEGEVOIVSTATOTFLAITSINGVLTWYVHGAGTTRIASPKGPVTOMYTN 92  
 Db 995 KTSMLGRDEREHSIVLTGTSRTSGMTGVNGVMTTFTGNSNARTLAGPVGPNCRWS 1054

QY 93 VDKDLVGWOAPOGSRSLTPTCGSSDLYLVTRHADVIVPVRGGDSRSLSPRISYLGK 152  
 Db 1055 PSDDVAVYPLPSGASCLPECKGTQSWCIRN--DGALCHGRSLKLVLDLPTLSDFRG 1112

QY 153 SSGGPLCLPAGHAGVIFRAAVSTRGV-----AKAVDFIPVES 189  
 Db 1113 SSGSPILCDGHHVGM--VSLHARGVKVTVGRVTPKWTLPKDS 1155

RESULT 13  
 T08839  
 C:Species: marmoset hepatitis GB virus A  
 C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C:Accession: T08839  
 R:Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J. Gen. Virol. 79, 41-45, 1998  
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A:Reference number: Z16486; MUID:98120818; PMID:9460920  
 A:Accession: T08839  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: genomic RNA  
 A:Residues: 1-2970 <ERK>  
 A:Cross-references: EMBL:AF023424; NID:q2828597; PIDN:AAC40501.1; PID:q2828598  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: polyprotein

Query Match 24.0%; Score 245; DB 2; Length 2970;  
 Best Local Similarity 27.8%; Pred. No. 9,2e-14;  
 Matches 62; Conservative 39; Mismatches 80; Indels 42; Gaps 6;

QY 3 KKGSVIVGRIN-----LSGDTAYAAQTRGCGCKTSHCRKNQVEGEVOIVS 52  
 Db 946 RRGDEVILIGLVGNWELPFGFVPTAPVYVHHHGKGFYVVKTSMTGWDTEHVGNNVYLG 1005

QY 53 TATOTFLAITSINGVLTWYVHGAGTTRIASPKGPVTOMYTNVDKLVGWOAPOGSRSLTPTC 112  
 Db 1006 TSTTRSMGTGVNGVMTTFTGNSNARTLAGQMPVNSRWNSASDDVAVPLPVGAKCLEPC 1065

QY 113 TCGSSDLYLVTRHADVIVPVRGGDSRSLLS-----PRPISYLGSGSGGFLCLCP 161  
 Db 1065 KCOPOGVVVI-----RND--GALCHGTLGRVLDLPAELCDFRGSGSGPILCD 1112

QY 162 AGHAGVIFRAAVSTRG-----VAKAVDFIPVESLETTMRSP 197  
 Db 1113 EGHAVGML-ISVLRGSRVGTGIRTKPWETLPREAITHTTEAPP 1154

## RESULT 14

H83144  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83144  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: H83144  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <STO>  
 A:Cross-references: GB:AE004818; GB:AE004091; NID:g9950200; PIDN:AAG07406.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4019  
 C:Superfamily: dedF protein

Query Match 8.4%; Score 85.5; DB 2; Length 209;  
 Best Local Similarity 27.9%; Pred. No. 1.7;  
 Matches 51; Conservative 16; Mismatches 61; Indels 55; Gaps 11;

QY 43 OYGEVQ-IVSTATOTFLAITSINGVL-----WTVYHGAGTTRIASPKGPVTOMYT 91  
 Db 29 QREVEVHFLISKAQLVMTETDVALPAKPOAMQAFLEYCGAAGQI-----RVFG 80

QY 92 NVDKDLVGWOAPOGSRSLTP-----CTCGSSDL-----YLVTRHADVIVPVRGGDS 137  
 Db 81 QND-----WVAPPASGSSAPNAWVICPSTGTLSAVATGACNLIERAADVALKER--- 131

QY 138 RGSLLSPR--PIS-----YLGSSGGPLCLPAGHAGVIFRAAVSTRGVAKAVDFIPVES 189  
 Db 132 RPLVLVPREAPFSSIHLENMLKLSNLGAVILPA--APGFYH---OPOSVELVDVFWVARI 186

QY 190 LET 192  
 Db 187 LNT 189

## RESULT 15

B71284  
 C:Species: Treponema pallidum subspp. pallidum (syphilis spirochete)  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 09-Dec-2002  
 C:Accession: B71284  
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gerson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; Moseley, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A:Reference number: A71250; MUID:98332770; PMID:9665876  
 A:Accession: B71284  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-398 <COL>

A:Cross-references: GB:AF001248; GB:AF000520; MID:g3323074; PIDN:AAC65740.1; PID:g332308  
A:Experimental source: strain Nichols  
C:Genetics: TP0773  
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryP

Query Match	8.4%	Score 85.5;	DB 2;	Length 398;
Best Local Similarity	23.0%	Pred. No. 3.7;		
Matches	44;	Conservative 29;	Mismatches 49;	Indels 69; Gaps 10;

  

Qy	59	LATSGVLM-----TVYHGAGTIRTIASPKGPV-----TQMY-----	90
Db	84	ITTEMVGVNMFLEPVPLEGGSGGAIIDARGYVLINTHVIEGASKIYLSLHDGSOYKATV	143
Qy	91	TNVDKD---LVGWAQPGSRSLTPTCTCGSSDLYLVTRHADV-----	128
Db	144	VGVDRENDLAVLFVSPPGAR-LTVIRFGSS-----RNLDVGQKVLAINPFGRLARTLT	196
Qy	129	-----IPVRRGD-SRGSLLSPRPISYLGSSGGLICPAGHAVGIFRAAVSTRGVA	179
Db	197	VGVSALARPIONKGSIRNNIQTDAAIN--PGNSGGPLDQTQRMIGINTVIYSTSGSS	254
Qy	180	KAVDF-IPVES	189
Db	255	SGVGFVAVPVD	265

Search completed: August 30, 2003, 19:20:31  
Job time : 17.2134 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 : Search time 9.75674 Seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKKKGSWIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	854.5	83.8	3011	1 POLG_HCV1	P26664 h genome po
2	848.5	83.2	3011	1 POLG_HCVH	P27958 h genome po
3	836.5	82.0	3010	1 POLG_HCVTW	P29846 h genome po
4	827.5	81.1	3010	1 POLG_HCVJT	Q00269 h genome po
5	823.5	80.7	3010	1 POLG_HCVBK	P26663 h genome po
6	823.5	80.7	3010	1 POLG_HCVJA	P26662 h genome po
7	675	66.2	3033	1 POLG_HCVJ8	P26661 h genome po
8	673	66.0	3033	1 POLG_HCVJ6	P26660 h genome po
9	87	8.5	321	1 HHOA_ARATH	Q9se17 arabidopsis
10	85.5	8.4	209	1 PAAD_PSEAE	Q9hx08 pseudomonas
11	83.5	8.2	437	1 DEGI_ARATH	O22609 arabidopsis
12	83	8.1	452	1 AAMP_HUMAN	Q13685 homo sapien
13	78.5	7.7	485	1 Y136_TREPA	O83172 treponema p
14	78.5	7.7	764	1 ICCR_DROME	O08180 drosophila
15	78	7.6	1165	1 POLGALV	P21414 gibbon ape
16	77.5	7.6	263	1 GRAK_MOUSE	O35205 mus musculus
17	76.5	7.5	323	1 VPRT_SMRVH	P21407 squirrel mo
18	76.5	7.5	333	1 MOSA_RHIME	Q07607 rhizobium m
19	76	7.5	401	1 FXHI_MOUSE	O88621 mus musculus
20	76	7.5	3411	1 POLG_YERV1	P03314 y genome po
21	76	7.5	3411	1 POLG_YERV2	P19901 y genome po
22	76	7.5	3414	1 POLG_TBVM	P14336 t genome po
23	75.5	7.4	248	1 TRY1_CHICK	Q90627 gallus gall
24	75.5	7.4	452	1 MLTD_ECOLI	P23931 escherichia
25	75.5	7.4	2269	1 WDR9_HUMAN	Q9ns16 homo sapien
26	75	7.4	467	1 NX1B_BOVIN	Q28142 bos taurus
27	74.5	7.3	248	1 GRAD_MOUSE	P11033 mus musculus
28	74.5	7.3	3414	1 POLG_LANVT	P29837 l genome po
29	74	7.3	911	1 TB1L_NEUTB	Q09056 neisseria m
30	74	7.3	973	1 VP18_HUMAN	Q9p253 homo sapien
31	74	7.3	3414	1 POLG_TBVM	Q01299 t genome po
32	73.5	7.2	248	1 TRY2_CHICK	Q90628 gallus gall
33	73.5	7.2	264	1 CTRL_HUMAN	P40313 homo sapien

RESULT 1  
POLG\_HCV1  
ID POLG\_HCV1 STANDARD; PRT: 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin) DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DE Hepatitis C virus (isolate 1) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Koit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;  
RT "Genetic organization and diversity of the hepatitis C virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GPNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: M62321; AAA45676.1; --  
CC PIR: A39166; GNVVC3  
CC FDB: 1AIV; 16-FEB-99.  
CC FDB: 1HEI; 25-NOV-98.  
CC MEROPS: S29.001; --  
CC MEROPS: U39.001; --  
CC InterPro: IPR001410; DEAD.  
CC InterPro: IPR002522; HCV\_capsid.

34 73.5 7.2 294 1 DPM1\_USTMA  
35 73.5 7.2 301 1 MCP\_BPF41  
36 73.5 7.2 443 1 FLII\_AQUAE  
37 73.5 7.2 706 1 TRPE\_HORSE  
38 73.5 7.2 1425 1 NPH4\_MOUSE  
39 73 7.2 478 1 MM03\_RABIT  
40 73 7.2 1530 1 NX1A\_BOVIN  
41 73 7.2 3412 1 POLG\_TBVM  
42 73 7.2 3415 1 POLG\_POWVL  
43 72.5 7.1 660 1 VST2\_HEVBU  
44 72.5 7.1 660 1 VST2\_HEVPA  
45 72.5 7.1 2499 1 MPRI\_BOVIN

#### ALIGNMENTS

DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR002518; HCV NS2.  
 DR InterPro: IPR004109; HCV NS3.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRP.  
 DR InterPro: IPR001650; Helicase.C.  
 DR InterPro: IPR007095; RNA\_pol\_DS-ps.  
 DR InterPro: IPR007094; RNA\_pol\_Psivir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase.C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Core protein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).  
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH\_BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5A99 CRC64;  
 Query Match 83.8%; Score 854.5; DB 1; Length 3011;  
 Best Local Similarity 82.8%; Pred. No. 2.5e-72;  
 Matches 169; Conservative 9; Mismatches 17; Indels 9; Gaps 1;  
 3 KGSVVIVGRIN-----LSGDTAYAQQTGEGGCKTSHTGRDNKQVEGEVQIVST 53

QY

Db 1005 RRGREILLGPADGKVKSGMRLAPITAYACOTRGLLCITSLTGRDNKQVEGEVQIVST 1064  
 QY 54 ATQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMYTNVDKDLGVQWAPQGSRLTPTCT 113  
 Db 1065 AAQTFLATSLNGVLTWVYHGAGTRTIASPKGPVTQMYTNVDQDLGVWAPQGSRLTPTCT 1124  
 QY 114 CGSSDLYLVTRHADVIPVRRRGRSGLLSRPISYLGKSSGGPLLCPAGHAGVIFRAAV 173  
 Db 1125 CGSSDLYLVTRHADVIPVRRRGRSGLLSRPISYLGKSSGGPLLCPAGHAGVIFRAAV 1184  
 QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN [1]  
 RP MEDLINE-92052256; PubMed-1659800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 [2]  
 RL X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RP MEDLINE-97331322; PubMed-9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RT Nat. Struct. Biol. 4:463-467(1997).  
 [3]  
 RL X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RP MEDLINE-98154321; PubMed-9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.





```
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;
QY 3 KGSVVIVGRIN-----LSGDTAYAQQTREGCCQKTSHTGRDKNQVEGEQIVST 53
Db 1005 RGQEIILGADGMVSKWRLAPITAYAAQTGRLGCIITSLTGRDKNQVEGEQIVST 1064
QY 54 ATOTFLATISVGLWTVYHGAGTRTTASPKGPVQTYTNVDKDLVQWAPQGSRSITPCT 113
Db 1065 ATOTFLATCINGCVTVYHGAGTRTTASPKGPVQTYTNVDKDLVQWAPQGSRSITPCT 1124
QY 114 CGSSDLYLVRHADVIPVRRGRSGSLSPRISYLVKSGSSGGLPCPAGHAVGIFRAAV 173
Db 1125 CGSSDLYLVRHADVIPVRRGRSGSLSPRISYLVKSGSSGGLPCPAGHAVGIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208
RESULT 3
POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92230206; PubMed=1314449;
RX Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RL mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84754; -; NOT_ANNOTATED_CDS.
CC PIR: A40244; GKNVTH.
CC PDB: 1N64; 25-FEB-03.
CC MEROPS: S29.001; -.
CC MEROPS: 039.001; -.
CC InterPro: IPR001410; DEAD.
```

```
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 131 CORE PROTEIN (POTENTIAL).
FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 2014 3010 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
```

Query Match 82.0%; Score 836.5; DB 1; Length 3010;  
Best Local Similarity 78.4%; Pred. No. 1.3e-70;  
Matches 160; Conservative 18; Mismatches 17; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGCQKTSHTGRDKNQVEGEVQIVST 53  
Db 1005 RRGREILLGPADSLGRGWRLLAPITAYAAQOTRGLFCIIITSLTGRDKNQVEGEVQIVST 1064  
Qy 54 ATQTFPLATSLVWLYHAGAGRTTASPGKPTOMYTNVDKDLVGHQAQPGSRSLTPCT 113  
Db 1065 ATOSFLATCLNGVCWTVYHAGSKTLAGPKPTOMYTNVDQDLVGHQAQPGSRSLTPCT 1124  
Qy 114 CGSSDLYLVTRHADVIPRRRGDSRGSLLSPRPISYLGSGGPGLLCPAGHAGVIFRAAV 173  
Db 1125 CGSSDLYLVTRHADVIPRRRGDSRGSLLSPRPISYLGSGGPGLLCPGSHVYVIFRAAV 1184  
Qy 174 STRGVAKNDPIPVSELETHMRSP 197  
Db 1185 CTRGVAKAVDFPVSEMETTHMRSP 1208

RESULT 4  
POLG\_HCVJT STANDARD; PRT: 3010 AA.  
AC Q00269;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66); (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus [isolate HC-JT] (HCV).  
OC Hepacivirus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OX NCBI\_TaxID=31642;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed-1318627;  
RA MEDLINE-92295714; PubMed-1318627;  
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
RT carrier: sequence variation within the same individual and among  
RT infected individuals."  
RL Virus Res. 23:39-53(1992).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC {RNA}(N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D11168; BAA01943.1; -  
CC PIR: A45573; A45573.  
CC PDB: 1A1Q; 25-MAR-98.  
CC PDB: 1JXP; 14-JAN-98.  
CC MEROPS: S29.001; -  
CC MEROPS: U39.001; -  
CC InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRp.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; helicase\_C; 1.  
DR Pfam: PF00998; Viral\_RdRp; 1.  
DR Pfam: PD186062; HCV\_NS1; 1.  
DR Pfam: PD186062; DEXDc; 1.  
DR SMART; SM00487; DEXDc; 1.  
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure. 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
FT INIT\_MET 1 1 CELLULAR AMINOPEPTIDASE.  
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).  
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).  
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
FT CHAIN 347 369 POTENTIAL.  
FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
FT SITE 1316 1319 DECH BOX.  
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 81.1%; Score 827.5; DB 1; Length 3010;  
Best Local Similarity 77.0%; Pred. No. 8.9e-70;  
Matches 157; Conservative 20; Mismatches 18; Indels 9; Gaps 1;  
Qy 3 KGSVVIVGRIN-----LSGDTAYAAQOTRGEQGCQKTSHTGRDKNQVEGEVQIVST 53

```

DB 1005 RRCRELLGPADSIQGGHRLAPITAYAQOTRGLGCVITSLTGRDNQVGEQVYST 1064
QY 54 ATQTFLATSLVLTVYHAGCTRIASPKGPTOMYTNVDKDLVGOAPOGSRSLTPTCT 113
DB 1065 ATQSFATCVNGCVTVFHCAGSKILAGPKGITOMYTNVDQDLVGHAPGARSILTPTCT 1124
QY 114 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRISYLVKSSGGPLLCPCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADYIPVRRRGDSRGLSPRISYLVKSSGGPLLCPCAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

```

## RESULT 5

```

ID POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed=8861916;
RA Love R.A., Farge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuebler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).

```

```

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M58335; AAA72945.1; .
CC PIR; A38465; GNMVTC.
CC PDB; 1A1Q; 25-MAR-98.
CC PDB; 1JXP; 14-JAN-98.
CC PDB; 1NS3; 08-APR-98.
CC PDB; 1C2P; 15-NOV-00.
CC PDB; 1GXS; 08-NOV-99.
CC PDB; 1GX6; 10-APR-02.
CC PDB; 1QTV; 26-JUN-00.
CC PDB; 80HM; 20-APR-99.
CC MEROPS; S29.001; .
CC
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV capsid.
CC InterPro; IPR002521; HCV core.
CC InterPro; IPR002519; HCV env.
CC InterPro; IPR002531; HCV NS1.
CC InterPro; IPR002518; HCV NS2.
CC InterPro; IPR004109; HCV NS3.
CC InterPro; IPR000745; HCV NS4A.
CC InterPro; IPR001490; HCV NS4B.
CC InterPro; IPR002868; HCV NS5A.
CC InterPro; IPR002166; HCV RdRp.
CC InterPro; IPR007095; RNA_pol_DS_Ps.
CC InterPro; IPR007094; RNA_pol_Psvir.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00998; Viral_RdRp; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXDC; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
CC 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.

```

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; F8422D5ECCFDFD9C CRC64;

Query Match 80.7%; Score 823.5; DB 1; Length 3010;  
Best Local Similarity 76.5%; Pred. No. 2.1e-69;  
Matches 156; Conservative 21; Mismatches 18; Indels 9; Gaps 1;

Oy 3 KKGSVIVIGRIN-----LSGDTAYAOOTRGECCQKTSHTGRDKNOVEGEVQIVST 53  
Db 1005 RRGKILLGPADSLGGRLLRLAPITAYSQOTRLLGCIITSLTGRDKNOVEGEVQIVST 1064  
Oy 54 ATOTFLATSLNGVLWYTHGAGRTIIASPKGPVTQMTYNDVKDLVQWQAPQGSRLTPCT 113  
Db 1065 ATQSFLATCVNGVCMYVTHGAGSKTLAAPKGPITQMTYNDVQDLVQWPKPPGARSLLTPCT 1124  
Oy 114 CGSSDLXLVTRHADVIPVRRRGRSGLSPRPISYILKSGSGGLPLCPAGHANGVIFRAAV 173  
Db 1125 CGSSDLXLVTRHADVIPVRRRGRSGLSPRPISYILKSGSGGLPLCPAGHANGVIFRAAV 1184  
Oy 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESMETTMRSP 1208

RESULT 6  
POLG\_HCVJA STANDARD; PRT: 3010 AA.  
ID POLG\_HCVJA  
AC F26662;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
NS5B (P66) (P70) (RNA-directed RNA polymerase) (HCV)  
Hepatitis C virus (isolate Japanese) (HCV)  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Hepacivirus.  
OX NCBI\_TaxID=11116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91088550; PubMed=2175903;  
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
Sugimura I., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from  
Japanese patients with non-A, non-B hepatitis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
RN [2]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,  
Ohkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
RL FEBS Lett. 280:325-328(1991).  
CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
precursor polyprotein, commonly with Asp or Glu in the P6  
position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
[RNA](N).  
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: D90208; BAAL4233.1;  
DR PIR: A39253; GNMVCGJ.  
DR HSSP: P26663; LXP.  
DR MEROPS: S29.001;  
DR MEROPS: U39.001;  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV\_capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV\_env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRP.





DR Pfam: PF00271: helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 224 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 430  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 FT SEQUENCE 3033 AA; 329165 MW; P957F5C1A273BE9E CRC64;  
 Query Match 66.0%; Score 673; DB 1: Length 3033;  
 Best Local Similarity 68.7%; Pred No. 3.7e-55;  
 Matches 123; Conservative 27; Mismatches 29; Indels 0; Gaps 0;  
 QY 19 TAYAOOTRGECQNTSHTGRDNQVEGEVOIVSTATOTFLATSIINGVLTMYVHGAGTRT 78  
 DB 1034 TAYAOOTRGELGTIVVSMGTGRDKTEQAGEIOVLSTVTSQSLGTTISGLVLTMYVHGAGNKT 1093  
 QY 79 IASPKGPVTOMTYNDKDLVGWAOPOGSRSLTPTCCGSSDLYLVRHADYIPVRRGDSR 138  
 DB 1094 LAGSRGPVTOMYSRREGDLGWSPPTGKSLPTCCGAVDLYLVRNADYIPARRRGDKR 1153  
 QY 139 GSLLSPRISYLUKSGGGLPLCLPAGHAYGIFRAAVSTRCYAKAVDFIPVRSLETTMRSP 197  
 DB 1154 GALLSPRLSTLUKSGSGPVLCPRHAGVGVFRAAVCSRGVAKSIDFIPVETLIVTRSP 1212  
 RESULT 9  
 HH0A\_ARATH STANDARD; PRT; 321 AA.  
 AC 09SEL7; 049507;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protease Hh0A, chloroplast precursor (EC 3.4.21.-).  
 GN Hh0A OR AT4G18370 OR F2B12.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OX eutrosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;  
 RT "Identification and characterization of the chloroplast RhoA protease,  
 RL a homolog to the bacterial periplasmic protease Hh0A.";  
 RN Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ansong W., Brandt P., Grivell L., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,  
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenot O., Quigley F., Clabaud J., Muendlein A., Felber R.,  
 RA Schnabl S., Hillier F., Schmidt W., Lecharny A., Aubourg S.,  
 RA Chedford T., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 RA Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RL thaliana.";  
 RN Nature 402:769-777(1999).  
 RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.  
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,  
 RA Kieselbach T.;  
 RT "The chloroplast lumen from Arabidopsis thaliana.";  
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.  
 CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
 CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
 CC gene model prediction. AT4G18370 and AT4G18375 were originally  
 CC fused into a single gene.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its



RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an

RP SEQUENCE OF 104-118.





Db 296 GSVDCQAKVYATTKVGVFRPETAASOPSLGEGESESNSVESL 341

```

RESULT 13
Y136_TREPA
ID Y136_TREPA STANDARD; PRT: 485 AA.
AC 083172:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical lipoprotein TP0136 precursor.
GN TP0136.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_Taxid=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodgson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388 (1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE001199; AAC65137.1; ALT_INIT.
DR TIGR: TP0136.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.
FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 164 178 GLY/SER-RICH.
FT DOMAIN 196 210 GLY/SER-RICH.
FT DOMAIN 253 267 GLY/SER-RICH.
FT DOMAIN 318 327 POLY-SER.
FT DOMAIN 444 447 POLY-SER.
SQ SEQUENCE 485 AA: 48984 MW: C7A4CEEDC7DC5CED CRC64;
Query Match 7.7% Score 78.5; DB 1: Length 485;
Best Local Similarity 23.4%; Pred. No. 6.9;
Matches 50; Conservative 16; Mismatches 77; Indels 71; Gaps 10;
QY 16 SGDITAYA-----QOTRGEQCKQKTSH-----TGRDKNQVEGVQIVSTATQTFPLATSI- 63
Db 54 AGSKLYATNGRLWEKELNGTSGWKVSSSVPTSDK-----KVMSTATDGNTEFLVACP 108
QY 64 -NGVLWTVYHGAG-----TFTIASPGQPVQTMVNDKDLVG-----NQAPQGSRLTPCT 113
Db 109 GTGVYKHCYKNGAGSSSTGTATSPSTETCSQAT-----LVGGTSKPFVLPVGGTNGNCG 164
QY 114 C-----GSSDLYLVTRHADVP-----VRRGDSRGLSLLSPRISYLK----- 151
Db 165 CGGGGGGSSSSSSSIIHLVPGGTGNGNCGCGGGGGGSSSSSIIHKIVKENTDQFL 224
QY 152 -----GSSGGLLCPAGHVG 167

```

Db 225 DMGEYVVTTKHLYTKNGSSSAGPAQCPCGGGGG 258

```

RESULT 14
ICCR_DROME
ID ICCR_DROME STANDARD; PRT: 764 AA.
AC Q08180:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Irregular chiasm C-roughnest protein precursor (IRREC protein).
GN RST.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94102535; PubMed=7503814;
RA Ramos R.G., Igloi G.L., Lichte B., Baumann U., Maier D.,
RA Schneider T., Brandstaetter J.H., Froehlich A., Fischbach K.-F.;
RT "The irregular chiasm C-roughnest locus of Drosophila, which affects
RT axonal projections and programmed cell death, encodes a novel
RT immunoglobulin-like protein.";
RL Genes Dev. 7:2533-2547 (1993).
CC -1- FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
CC RETINA.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
CC -1- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
CC IN LATE LARVAL AND PUPAL STAGES.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z21641; CAA79756.1;
DR EMBL: L11040; AAA16632.1;
DR PIR: A49448; A49448.
DR FlyBase: FBgn003285; rst.
DR GO: GO:0016202; P:regulation of myogenesis; IMP.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003598; Ig-C2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig; 4.
DR SMART: SM00408; Igc2; 1.
DR PROSITE: PS50835; IG_LIKE; 5.
DR Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat;
KW Cell adhesion.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 764 IRREGULAR CHIASM C-ROUGHNEST PROTEIN.
FT DOMAIN 20 533 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 534 556 POTENTIAL.
FT DOMAIN 557 764 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 123 IG-LIKE C2-TYPE 1.
FT DOMAIN 117 230 IG-LIKE C2-TYPE 2.
FT DOMAIN 245 261 GLY-RICH.
FT DOMAIN 237 343 IG-LIKE C2-TYPE 3.
FT DOMAIN 346 419 IG-LIKE C2-TYPE 4.
FT DOMAIN 430 530 IG-LIKE C2-TYPE 5.
FT DOMAIN 637 660 GLN-RICH (OPA-REPEAT).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```
FT CARBOHYD 507 507 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 764 AA; 82947 MW; 262225D2B2A1C181 CRC64;

Query Match
Best Local Similarity 20.3%; Score 78.5; DB 1; Length 764;
Matches 38; Conservative 27; Mismatches 69; Indels 53; Gaps 7;

QY 27 GEQGCOKTSHTGRDKNOVEGEVQIVSTATQTFPLATSIING-----VLWTVYHGAG 75
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 504 KYNCTVNDYGN-----VAEIQLOAKKSVLLMTIVGSIWVAFLLVLTILVVYIKCK 559
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 76 TRTIASP-----KGPVTOMTINVDKDLGVQWQAPQSGRS----- 108
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 560 KRTKLPADVISEHQITKNGVSCLEPGDRTSNISDLKVDISGGYVPYGYSTHYSPPP 619
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 109 --LTPTCTGCS--SDLYLVTRHADVIPIVRRRGD-----SRGSLSPRPISYLYKSGSGGPLLC 160
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 620 QYLTCSTKSNSSITMNNHQNLOLOQQQQOSSHQHHHTQTTLPTFLINSSGSL-- 677
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 161 PAGHVG 167
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 678 -TGSIIIG 683

RESULT 15
POL_GALV
ID POL_GALV STANDARD; PRT: 1165 AA.
AC P21414;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pol polyprotein [Contains: Protease (EC 3.4.23.-); Reverse
DE transcriptase (EC 2.7.7.49); Endonuclease].
GN POL.
OS Gibbon ape leukemia virus.
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11840;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=90051069; PubMed=2683360;
RA Delassus S., Sonigo P., Wain-Hobson S.;
RT "Genetic organization of gibbon ape leukemia virus.";
RL Virology 173:205-213(1989).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + {DNA}(n).
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M26927; AAA46810.1; -
CC PIR: B32595; GNJLGL.
CC HSP: P03355; IMML.
CC MEROPS: A02.008; -.
CC InterPro: IPR001995; Aspprotease_rtrv.
CC InterPro: IPR001969; Aspprotease_site.
CC InterPro: IPR002156; RNaseH.
CC InterPro: IPR001584; Rve.
CC InterPro: IPR000477; RVTse.
CC Pfam: PF00075; rnaseH; 1.
CC Pfam: PF00665; rve; 1.
CC Pfam: PF00077; rvp; 1.
CC Pfam: PF00078; rvt; 1.
CC PROSITE: PS00141; ASP_PROTEASE; 1.
CC PROSITE: PSS0175; ASP_PROT_RETROV; 1.
KW Hydrolase; Transferase; RNA-directed DNA polymerase;
KW Aspartyl protease; Endonuclease; Polyprotein.
```

```
FT CHAIN 1 103 PROTEASE.
FT ACT SITE 27 27 BY SIMILARITY.
SQ SEQUENCE 1165 AA; 129886 MW; 8B7AFD54812B7E1A CRC64;

Query Match
Best Local Similarity 22.6%; Score 78; DB 1; Length 1165;
Matches 53; Conservative 33; Mismatches 85; Indels 64; Gaps 11;

QY 3 KGSVVIVGRINLSGDTAVAAQQTGRGQCGQK-TSHTG-----RDKNOVEGEV 48
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 354 KKGTKLLQELSLG---YRVSAAKQALCOREVTVLYLLKEGKRWLTTPARKATVTKIPV 410
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 49 QIVSTATQTFPLATSIINGVLWTVYHGAGTRTITASPKGPVT-----QMTYNVDK 95
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 411 PTTTPROVREFLGTAFCRLWI-----PCFASLAAPLYPLTKESIPFIWTEHQQAEDHIKK 466
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 96 DLYGWAQ---PQGSRSITPTCTCGSSDLYLVTRHADVIPIVRRRGDSRGL---LSP--RPI 147
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 467 ALLSAPALALPDLTKPFT-----LYI-----DERAGVARGVLTQTLGPMRREV 509
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 148 SYLKG-----SSGGLPLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMRSP 197
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 510 AYLSKKLDPVASGWPCLKAAVAALLLLKDDADKLTLGQNTVVIASHLSLESIVRQP 564
```

Search completed: August 30, 2003, 19:13:49  
Job time : 10.7567 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.5921 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKRKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organalle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872.5	85.5	4040	12 Q9IFH8	Q9ifh8 mucosal dis
2	858.5	84.2	3011	12 Q36579	Q36579 hepatitis c
3	854.5	83.8	2436	12 Q81756	Q81756 hepatitis c
4	854.5	83.8	3011	12 Q91FE5	Q91fe5 hepatitis c
5	854.5	83.8	3011	12 Q9ELS8	Q9els8 hepatitis c
6	853.5	83.7	3011	12 Q03463	Q03463 hepatitis c
7	851.5	83.5	3011	12 Q36608	Q36608 hepatitis c
8	851.5	83.5	3015	12 Q9PWX5	Q9pwx5 hepatitis c
9	851.5	83.5	3015	12 Q9PWX9	Q9pwx9 hepatitis c
10	849	83.2	181	12 Q91RR8	Q91rr8 hepatitis c
11	849	83.2	181	12 Q91RT5	Q91rt5 hepatitis c
12	847	83.0	181	12 Q91RR5	Q91rr5 hepatitis c
13	847	83.0	181	12 Q91RR2	Q91rr2 hepatitis c
14	847	83.0	181	12 Q91RT9	Q91rt9 hepatitis c
15	846	82.9	181	12 Q91RR3	Q91rr3 hepatitis c
16	846	82.9	181	12 Q91RR4	Q91rr4 hepatitis c

17	846	82.9	181	12 Q91RS1	Q91rs1 hepatitis c
18	846	82.9	181	12 Q91RQ8	Q91rq8 hepatitis c
19	846	82.9	181	12 Q91RT1	Q91rt1 hepatitis c
20	846	82.9	181	12 Q91RR0	Q91rr0 hepatitis c
21	845.5	82.9	3011	12 Q36609	Q36609 hepatitis c
22	844	82.7	181	12 Q91RR6	Q91rr6 hepatitis c
23	844	82.7	181	12 Q91RS9	Q91rs9 hepatitis c
24	843	82.6	181	12 Q91RS3	Q91rs3 hepatitis c
25	842.5	82.6	3011	12 Q9DIT6	Q9dit6 hepatitis c
26	842	82.5	181	12 Q91RT4	Q91rt4 hepatitis c
27	842	82.5	181	12 Q91RS8	Q91rs8 hepatitis c
28	842	82.5	181	12 Q91RT3	Q91rt3 hepatitis c
29	842	82.5	181	12 Q91RS5	Q91rs5 hepatitis c
30	842	82.5	181	12 Q91RS7	Q91rs7 hepatitis c
31	842	82.5	181	12 Q91RT0	Q91rt0 hepatitis c
32	842	82.5	181	12 Q91RS2	Q91rs2 hepatitis c
33	841	82.5	181	12 Q91RS6	Q91rs6 hepatitis c
34	840.5	82.4	3010	12 Q9OP61	Q9op61 hepatitis c
35	840	82.4	181	12 Q91RS4	Q91rs4 hepatitis c
36	839.5	82.3	3010	12 Q68533	Q68533 hepatitis c
37	839	82.3	181	12 Q91RR7	Q91rr7 hepatitis c
38	839	82.3	181	12 Q91RT6	Q91rt6 hepatitis c
39	839	82.3	3011	12 Q36610	Q36610 hepatitis c
40	838	82.2	181	12 Q91RT8	Q91rt8 hepatitis c
41	837.5	82.1	361	12 Q70818	Q70818 hepatitis c
42	837.5	82.1	361	12 Q70817	Q70817 hepatitis c
43	837	82.1	181	12 Q91RR9	Q91rr9 hepatitis c
44	836.5	82.0	3010	12 Q9DTE2	Q9dte2 hepatitis c
45	836.5	82.0	3010	12 Q99AU2	Q99au2 hepatitis c

#### ALIGNMENTS

#### RESULT 1

Q9IFH8 ID Q9IFH8 PRELIMINARY; PRT; 4040 AA.  
AC Q9IFH8; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Genome polypeptide.  
OS Mucosal disease virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OX Pestivirus.  
NCBI\_TaxID=11099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20323484; PubMed-10864644;  
RA Lal V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,  
RA Denis R.O., Hong Z., Lau J.Y.;  
RT \*Generation and characterization of a hepatitis C virus NS3 protease-  
dependent bovine viral diarrhea virus.\*;  
RL J. Virol. 74:6339-6347(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Lal V.C.H., Hong Z.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF268278; AAF82566.1; -;  
DR HSSP; P26663; 1JXP.  
DR MEROPS; S31.001; -;  
DR InterPro; IPR000280; CDvir\_endptsep80.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR004109; HCV\_NS3.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase.C.  
DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF002907; HCV\_NS3; 1.  
DR Pfam; PF00271; Helicase.C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.

DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_Core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NSI; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00998; Viral\_rdrP; 1.  
DR ProDom; PD186062; HCV\_NSI; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS05007; RDRP\_POSITIVE; 1.  
DR PROSITE; PS05021; RDRP\_VIRAL; 1.  
DR KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydrolase; Nonstructural protein; Polyprotein;  
KW RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3011 AA; 377182 MW; E2E0EE809C63CB9 CRC64;

Query Match 84.2%; Score 858.5; DB 12; Length 3011;  
Best Local Similarity 82.8%; Pred. No. 2.2e-73;  
Matches 169; Conservative % 10; Mismatches 16; Indels 9; Gaps 1;

QY 3 KKGSVVIVGRIN-----LSGDYAAQOTRGEQGCGCKTSHTGRDKNVEGEVOIVST 53  
DB :|||:|||||:|\_|:|||||: || |:|||:|||||:|||||:  
1005 RRGEILLGPADGHWKVGWELLAPITAYAAQTTRGLLGCIITSLTGRDNKEGEVQIVST 1064

QY 54 ATQTFLATISINGLVTVYHGAGTRTIASPKGPVTQMTYNVDKLDYGWAOPQGSRLTPCT 113  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1065 ATQTFLATINCINGVTVYHGAGTRTIASPKGPVIQMYTNWDQLVGGWPAPOGSRSLTPCT 1124

QY 114 CGSSDLXLYLTHADVIPVRERGDSRGSLSPRISYLKSGSGGPLLCPGAHAVGIFFRAAV 173  
DB :|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1125 CGSSDLXLYLTHADVIPVRERGDSRGSLSPRISYLKSGSGGPLLCPGAHAVGIFFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
DB :|||||:|||||:|||||:|||||:|||||:  
1185 CTBGVAKAVDFIPVENLETTMRSP 1208

RESULT 3  
O81756 PRELIMINARY; PRT; 2436 AA.

ID O81756 AC  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Genome Polyprotein (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
RN NCBI\_TaxID=111103;  
RX [1]  
RP CHOO Q.-L., Richman K., Han J.;  
RA "The nucleotide sequence of the Hepatitis C viral genome.";  
RT Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.  
RL EMBL: W32084; AAA45677.1; -

DR HSP; P27958; IAIv.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR002531; HCV\_NSI.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR004109; HCV\_NS3.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_rdrP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVlr.  
DR Pfam; PF01560; HCV\_NSI; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.



```

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3E13F CRC64;

Query Match      83.8%; Score 854.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 5.3e-73;
Matches 168; Conservative 11; Mismatches 16; Indels 9; Gaps 1;

QY 3 KGSWVIVGRN-----LSGDTAYAQOTRGEQCCQKTSHTGRDKNQVGEVQIVST 53
DB 1005 RGQEILGPADGMVSKGWRLAPITAYAQOTRGLGCIITSLTGRDKNQVGEVQIVST 1064

QY 54 ATQTFLATISNGVLTWYHVGAGTITIASPKGPVTOMYTNVDKLVGWAQPGSRSITPCT 113
DB 1065 ATQTFLATCINGVLTWYHVGAGTITIASPKGPVQMTNTVDQDLVGPAPGSGRSITPCT 1124

QY 114 CGSDDLVLVTRHADYIPVRRGDSRGSLLSPRPISYLGKSGGPGLLCPAGHAYGIFRAAV 173
DB 1125 CGSDDLVLVTRHADYIPVRRGDSRGSLLSPRPISYLGKSGGPGLLCPAGHAYGLFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETMRSP 1208

RESULT 6
Q03463
ID Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;

```

```

RT *The 5'-terminal sequence of the hepatitis C virus genome.*;
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT *Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.*;
RN J. Gen. Virol. 72:2697-2704(1991).
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT *Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.*;
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT *Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.*;
RL Virology 200:246-255(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL: D10749; BAA01582.1; -.
DR HSP; P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

```

```
Query Match      83.7%; Score 853.5; DB 12; Length 3011;
Best Local Similarity 82.8%; Pred. No. 6.6e-73;
Matches 169; Conservative 8; Mismatches 18; Indels 9; Gaps 1;

QY      3 KKGWVIVGRIN-----LSGDTAYAQQTRGEGCGCKTSHTGRDKNQVEGEVQIVST 53
DB      1005 RRGQEILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLSGRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATISNGVLVTVHAGCTRTIASPKGPVTQMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB      1065 AAGTFLATCINGVCTVYHAGCTRTIASPKGPVIOMYTNVDODLVGMQAPQARSLTPTCT 1124
QY      114 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPCAGHAGVIFRAAV 173
DB      1125 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPCAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608          PRELIMINARY:      PRT: 3011 AA.
AC      O36608;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus strain H77.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=63746;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=H77.
RX      MEDLINE=97385173; PubMed=9238047;
RA      Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT      "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT      are infectious when directly transfected into the liver of a
RT      chimpanzee.";
RL      Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA (BY SIMILARITY).
DR      EMBL: AF011751; AAB67036.1;
DR      HSSP: P27958; 1HEI
DR      InterPro: IPR001410; DEAD.
DR      InterPro: IPR002522; HCV_capsid.
DR      InterPro: IPR002521; HCV_core.
DR      InterPro: IPR002519; HCV_env.
DR      InterPro: IPR002531; HCV_NS1.
DR      InterPro: IPR002518; HCV_NS2.
DR      InterPro: IPR002519; HCV_NS3.
DR      InterPro: IPR000745; HCV_NS4a.
DR      InterPro: IPR001490; HCV_NS4b.
DR      InterPro: IPR002868; HCV_NS5a.
DR      InterPro: IPR002166; HCV_RdRp.
DR      InterPro: IPR001650; Helicase_C.
DR      InterPro: IPR007095; RNA_pol_DS_PS.
DR      InterPro: IPR007094; RNA_pol_PSVir.
DR      Pfam: PF01543; HCV_capsid; 1.
DR      Pfam: PF01542; HCV_core; 1.
DR      Pfam: PF01539; HCV_env; 1.
DR      Pfam: PF01560; HCV_NS1; 1.
DR      Pfam: PF01538; HCV_NS2; 1.
DR      Pfam: PF02907; HCV_NS3; 1.
DR      Pfam: PF01006; HCV_NS4a; 1.
DR      Pfam: PF01006; HCV_NS4b; 1.
DR      Pfam: PF01001; HCV_NS4b; 1.
DR      Pfam: PF01506; HCV_NS5a; 1.
DR      Pfam: PF00271; helicase_C; 1.
```

```
DR      Pfam: PF00998; Viral_RdRp; 1.
DR      PRODOM: PD186062; HCV_NS1; 1.
DR      SMART: SM00487; DEXdc; 1.
DR      PROSITE: PS05057; RDRP_POSITIVE; 1.
DR      PROSITE: PS05052; RDRP_VIRAL; 1.
KW      ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW      Hydrolase; Nonstructural protein; Polyprotein;
KW      RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3011 AA; 327112 MW; 0B75E6881CB5C198 CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY      3 KKGWVIVGRIN-----LSGDTAYAQQTRGEGCGCKTSHTGRDKNQVEGEVQIVST 53
DB      1005 RRGQEILLGPADGMVSKGWRLLAPITAYAAQTRGLGCIITSLSGRDKNQVEGEVQIVST 1064
QY      54 ATQTFLATISNGVLVTVHAGCTRTIASPKGPVTQMYTNVDKDLVGMQAPQGSRLTPTCT 113
DB      1065 ATQTFLATCINGVCTVYHAGCTRTIASPKGPVTQMYTNVDQDLVGMQAPQGSRLTPTCT 1124
QY      114 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPCAGHAGVIFRAAV 173
DB      1125 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPCAGHAGVIFRAAV 1184
QY      174 STRGVAKAVDFIPVESLETTMRSP 197
DB      1185 CTRGVAKAVDFIPVENLGTMRSP 1208

RESULT 8
Q9PMX5          PRELIMINARY:      PRT: 3015 AA.
AC      Q9PMX5;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99420396; PubMed=10489358;
RA      Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT      "Hepatitis C virus: an infectious molecular clone of a second major
RT      genotype (2a) and lack of viability of intertypic 1a and 2a
RT      chimeras.";
RL      Virology 262:250-263(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Bukh J.;
RL      Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA (BY SIMILARITY).
DR      EMBL: AF177040; AAF01182.1;
DR      EMBL: AF177038; AAF01180.1;
DR      HSSP: P27958; 1HEI.
DR      InterPro: IPR001410; DEAD.
DR      InterPro: IPR002521; HCV_capsid.
DR      InterPro: IPR002521; HCV_core.
DR      InterPro: IPR002519; HCV_env.
DR      InterPro: IPR002531; HCV_NS1.
DR      InterPro: IPR002518; HCV_NS2.
DR      InterPro: IPR000745; HCV_NS3.
DR      InterPro: IPR000745; HCV_NS4a.
DR      InterPro: IPR001490; HCV_NS4b.
DR      InterPro: IPR002868; HCV_NS5a.
DR      InterPro: IPR002166; HCV_RdRp.
```



```

DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002129; Pyridoxal dec.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; 87023BCIF190863A CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGWVIVGRIN-----LSGDTAYAQTRGQCCKTSHTGRDNQVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLLAPITAYAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATQFLATSIINGVLTVYHGAGTRTIASPKGPVTOMTYNDKDLVGNQAPQGSRLTPTCT 113
DB 1069 ATQFLATCINGVCTVYHGAGTRTIASPKGPVTOMTYNDKDLVGNQAPQGSRLTPTCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGRSGSLLSPRPISYLGKSGGGLPCPAGHVGIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGRSGSLLSPRPISYLGKSGGGLPCPAGHVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 9
Q9PMU9
ID Q9PMU9 PRELIMINARY; PRT: 3015 AA.
AC Q9PMU9
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

SEQUENCE FROM N.A.
RP MEDLINE-99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.";
RL Virology 262:250-263(1999).
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF177039; AAF01181.1; -.
DR EMBL; AF177037; AAF01179.1; -.
DR HSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002129; Pyridoxal dec.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match      83.5%; Score 851.5; DB 12; Length 3015;
Best Local Similarity 82.4%; Pred. No. 1e-72;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KKGWVIVGRIN-----LSGDTAYAQTRGQCCKTSHTGRDNQVEGEVQIVST 53
DB 1009 RGQEILLGPADGMVSKGWRLLAPITAYAQTRGLGCIITSLTGRDNQVEGEVQIVST 1068

QY 54 ATQFLATSIINGVLTVYHGAGTRTIASPKGPVTOMTYNDKDLVGNQAPQGSRLTPTCT 113
DB 1069 ATQFLATCINGVCTVYHGAGTRTIASPKGPVTOMTYNDKDLVGNQAPQGSRLTPTCT 1128

QY 114 CGSSDLVLTTRHADVIPVRRGRSGSLLSPRPISYLGKSGGGLPCPAGHVGIFRAAV 173
DB 1129 CGSSDLVLTTRHADVIPVRRGRSGSLLSPRPISYLGKSGGGLPCPAGHVGIFRAAV 1188

QY 174 STRGVAKAVDFIPVESLETMRSP 197
DB 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 10
Q91RR8
ID Q91RR8 PRELIMINARY; PRT: 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE NS3 protease (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.1Y;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369235; AAK54560.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.2%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 4.6e-74;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAQOTRGEQCGCKTSHTGRDKNQVEGEQIVSTATQTFLATISINGVLTWYHAGTGT 78
DB  4 TAYAQOTRGLGCIITSLTGRDKNQVEGEQIVSTAAQTFLATCINGVCTWYHAGTGT 63
QY  79 IASPKGPVTQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 196
DB  124 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETMTMS 181

RESULT 11
Q91RT5
ID  Q91RT5 PRELIMINARY; PRT; 181 AA.
AC  Q91RT5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.4;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369218; AAK54543.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match      83.2%; Score 849; DB 12; Length 181;
Best Local Similarity 93.3%; Pred. No. 4.6e-74;
Matches 166; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAQOTRGEQCGCKTSHTGRDKNQVEGEQIVSTATQTFLATISINGVLTWYHAGTGT 78
DB  4 TAYAQOTRGLGCIITSLTGRDKNQVEGEQIVSTAAQTFLATCINGVCTWYHAGTGT 63
QY  79 IASPKGPVTQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 196
DB  124 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETMTMS 181

RESULT 12
Q91RR5
ID  Q91RR5 PRELIMINARY; PRT; 181 AA.
AC  Q91RR5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.30;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369238; AAK54563.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match      83.0%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 7.2e-74;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAQOTRGEQCGCKTSHTGRDKNQVEGEQIVSTATQTFLATISINGVLTWYHAGTGT 78
DB  4 TAYAQOTRGLGCIITSLTGRDKNQVEGEQIVSTAAQTFLATCINGVCTWYHAGTGT 63
QY  79 IASPKGPVTQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 196
DB  124 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 181

RESULT 13
Q91RR2
ID  Q91RR2 PRELIMINARY; PRT; 181 AA.
AC  Q91RR2;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.4V;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369241; AAK54566.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```

```

DB  124 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVENLETMTMS 181

RESULT 12
Q91RR5
ID  Q91RR5 PRELIMINARY; PRT; 181 AA.
AC  Q91RR5;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.30;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369238; AAK54563.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match      83.0%; Score 847; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 7.2e-74;
Matches 165; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY  19 TAYAQOTRGEQCGCKTSHTGRDKNQVEGEQIVSTATQTFLATISINGVLTWYHAGTGT 78
DB  4 TAYAQOTRGLGCIITSLTGRDKNQVEGEQIVSTAAQTFLATCINGVCTWYHAGTGT 63
QY  79 IASPKGPVTQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 138
DB  64 IASPKGPVIQMTYNDKDLVQWAPQGSRLTPTCGSSDLYLVTRHADVIPVRRGDSR 123
QY  139 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 196
DB  124 GSLLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETMTMS 181

RESULT 13
Q91RR2
ID  Q91RR2 PRELIMINARY; PRT; 181 AA.
AC  Q91RR2;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  NS3 protease (Fragment).
OS  Hepatitis C virus.
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX  NCBI_TaxID=11103;
RN  SEQUENCE FROM N.A.
RP  STRAIN-Pt.4V;
RA  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  *Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.*;
DR  EMBL; AF369241; AAK54566.1; -.
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3; 1.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run On: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-20  
Perfect score: 1020  
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-O=/cn2\_1/vsPTO\_spool/US09965594/runat\_29082003\_151919\_28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCI=0 -LPOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.@CGN\_1\_1.14686.@runat\_29082003\_151919\_28310 -NCPU=3  
-NO\_MMALP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:1  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	892.5	87.5	12734	6	ARI179057	ARI179057 Sequence
2	885.5	86.8	1998	6	ARI145264	ARI145264 Sequence
3	882.5	86.5	1998	6	ARI145268	ARI145268 Sequence
4	881.5	86.4	1998	6	ARI145263	ARI145263 Sequence
5	878.5	86.1	651	6	ARI145254	ARI145254 Sequence
6	878.5	86.1	1998	6	ARI145267	ARI145267 Sequence
7	877.5	86.0	1998	6	ARI145262	ARI145262 Sequence
8	875.5	85.8	651	6	ARI145258	ARI145258 Sequence
9	874.5	85.7	651	6	ARI145253	ARI145253 Sequence
10	874.5	85.7	1998	6	ARI145266	ARI145266 Sequence
11	873.5	85.6	1998	6	ARI145261	ARI145261 Sequence
12	873.5	85.6	2016	6	ARI145269	ARI145269 Sequence
13	872.5	85.5	12734	14	AF268278	AF268278 Pestivirus
14	871.5	85.4	651	6	ARI145257	ARI145257 Sequence
15	870.5	85.3	651	6	ARI145252	ARI145252 Sequence
16	870.5	85.3	1998	6	ARI145265	ARI145265 Sequence
17	870.5	85.3	2016	6	ARI145270	ARI145270 Sequence
18	870	85.3	648	6	ARI145274	ARI145274 Sequence
19	868	85.1	648	6	ARI145272	ARI145272 Sequence
20	867.5	85.0	651	6	ARI145256	ARI145256 Sequence
21	867.5	85.0	651	6	ARI145260	ARI145260 Sequence
22	866.5	85.0	651	6	ARI145251	ARI145251 Sequence
23	866	84.9	648	6	ARI145273	ARI145273 Sequence
24	864	84.7	648	6	ARI145271	ARI145271 Sequence
25	863.5	84.7	651	6	ARI145255	ARI145255 Sequence
26	863.5	84.7	651	6	ARI145259	ARI145259 Sequence
27	861	84.4	8157	6	ARI127810	ARI127810 Sequence
28	861	84.4	8157	6	BD081911	BD081911 Hepatitis
29	859	84.2	1932	6	ARI127809	ARI127809 Sequence
30	859	84.2	1932	6	BD081910	BD081910 Hepatitis
31	858.5	84.2	9646	6	BD110828	BD110828 Sequence
32	858.5	84.2	9646	6	BD069982	BD069982 Functional
33	858.5	84.2	9646	14	AF009606	AF009606 Hepatitis
34	858.5	84.2	12980	6	ARI110831	ARI110831 Sequence
35	858.5	84.2	12980	6	BD069985	BD069985 Functional
36	854.5	83.8	5360	6	ARI118686	ARI118686 Sequence
37	854.5	83.8	5360	6	I06434	I06434 Sequence 48
38	854.5	83.8	5360	6	I09328	I09328 Sequence 8
39	854.5	83.8	6785	6	ARI118692	ARI118692 Sequence
40	854.5	83.8	6785	6	I06440	I06440 Sequence 54
41	854.5	83.8	6785	6	I09329	I09329 Sequence 10
42	854.5	83.8	7310	6	ARI118696	ARI118696 Sequence
43	854.5	83.8	7310	6	I09331	I09331 Sequence 15
44	854.5	83.8	7310	14	HPCPOLYP	M32084 Hepatitis C
45	854.5	83.8	8316	6	ARI118703	ARI118703 Sequence

ALIGNMENTS

RESULT 1

AR179057  
LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6326137.  
ACCESSION AR179057  
VERSION AR179057.1 GI:20220612  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 12734)  
AUTHORS Hong, Z., Lai, V.C.H. and Lau, J.Y.N.  
TITLE Hepatitis C virus protease-dependent chimeric pestivirus  
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;  
FEATURES  
Location/Qualifiers  
source 1..12734  
BASE COUNT 4032 a 2604 c 3295 g 2803 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 12734  
Score: 892.50  
Length: 12734  
Matches: 177  
Percent Similarity: 92.82%  
Conservative: 4  
Best Local Similarity: 90.77%  
Mismatches: 11  
Query Match: 87.50%  
Indels: 3  
Gaps: 6  
DB: 1

US-09-965-594-20 (1-197) x AR179057 (1-12734)  
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
DB 413 GGTAGTGTGTTATTTGGTAGAATTGTTTATCTGTTAGTGTGTATATACACGGCTAC 472  
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyValAspLys 41  
DB 473 GCGGACAGACGAGGAGGCTCTAGGCTGTAGATCACCAGCTGACTGCGCGGACAAA 532  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 533 AACCAAGTGGAGGTGAGCTCCAGATGCTGAACCTGCTACCCAAACCTTCTGGCAACG 592  
QY 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 593 TGCATCAATGGGTATGCTGGACGTCTACCCAGGCGCGGACGAGGACATCGCATCA 652  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 653 CCCAAGGCTCTGTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGCTGCC 712  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 713 GCTCCTCAAGGTTCGGCTCATTCACACCTGACCTGCGGCTCTCGGACCTTTACCTG 772  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
DB 773 GTTACGAGGACGCGGACGTCTATCCCTGCGCGGCGGAGGTATAGCAGGGGTAGCCTG 832  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
DB 833 CTTTCGCGCGCGCGCATTTCTTACTAAAGGCTCTCGGGGGTTCGCTGTTGCGCCC 892  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 893 GCGGACACGCGCTGGGCTATTACAGGCGCGGCTGTCACCCGTGGAGTGGCCAAAGCG 952  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
DB 953 GTGGACTTTATCCCTGTGGAGAACCTAGACAGCAACCATGAGATCC 997

RESULT 2  
AR145264  
LOCUS AR145264 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 105 from patent US 6211338.

AR145264  
VERSION AR145264.1 GI:15107131  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1998)  
AUTHORS Malcolm, B.A., Taremi, S. Shane, Weber, P.C. and Yao, N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 105 03-APR-2001;  
FEATURES  
Location/Qualifiers  
source 1..1998  
BASE COUNT 411 a 595 c 569 g 423 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1998  
Score: 885.50  
Length: 1998  
Matches: 168  
Percent Similarity: 93.37%  
Conservative: 15  
Best Local Similarity: 85.71%  
Mismatches: 10  
Query Match: 86.81%  
Indels: 3  
Gaps: 1  
DB: 1

US-09-965-594-20 (1-197) x AR145264 (1-1998)  
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
DB 64 GGTCTGTGTTATTTGGTAGAATTATTTATCTGTTAGTGTGTATATACACGGCTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyValAspLys 41  
DB 124 TCCCAACAGACGCGGGCTACTTGTGTGCAAGAAGACTAGCTTACAGGCGGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
DB 184 AACCAAGTGGAGGTGAGCTCCAGTGTGTTCCACCGCAACACAACTCTCTCGCGCAC 243  
QY 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
DB 244 TGGCTCAAGCGCTGTGTGGACCTTTACCATGGTGTGCTCAAGACCTTAGCCGGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
DB 304 CCNAAAGGGCCCAATCACCCAGATGTACACTAATGTGGACAGGACCTCTCGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
DB 364 GCGCGCGCGGGCGGTCTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
DB 424 GTCACGAGACATGCTGACGCTCATTCGGTGCCTGCGCGGCGGACAGTAGGGGAGCCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
DB 484 CTCCTCCCGCAGCGCTCTCTCTACTTGAAGGCTCTTCGGGTGGTCCACTGCTGCGCCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
DB 544 TCGGGGACGCTGTGGGCATCTTCCGGGCTGCGGTATGCACCGCGGGGTTTGGGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
DB 604 GTGGACTTTTGTCCCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

RESULT 3  
AR145268  
LOCUS AR145268 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 109 from patent US 6211338.  
ACCESSION AR145268  
VERSION AR145268.1 GI:15107135

## KEYWORDS

Unknown.  
Unclassified.

## SOURCE

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolm, B.A., Taremi, S.Shane., Weber, P.C. and Yao, N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 109 03-APR-2001;

FEATURES Location/Qualifiers

source 1..1998

BASE COUNT 411 a 595 c 569 g 423 t

## ORIGIN

Alignment Scores:  
Pred. No.: 1,25e-65 Length: 1998  
Score: 882.50 Matches: 167  
Percent Similarity: 93.37% Conservatives: 16  
Best Local Similarity: 85.20% Mismatches: 10  
Query Match: 86.52% Indels: 3  
DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145268 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 64 GGTCTCTGTTATTTGTTAGTAATATTATTTATCTGCTAGTATCATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
Db 124 TCCACACAGACGGGGGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 184 AACCAAGTTCGAGGAGAGGTTTCAGTGGTTTCCACCGCAACACAACTCTTCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 244 TGCCTCAACGGCGTGTGTGGACCGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
Db 304 CCAAGGGGCGCAATCACCAGATGTACACTAATGTGACCAAGACCTCGTGGCGTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 364 GCGCCCCCGGGGCGCTTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 424 GTACAGAGACATGCTGACGTCAATTCGGGTGCGCGGGCGGCGAGTAGGGGAGCGTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 484 CTCTCCCCAGGCGCTGCTCTTCTTGAAGGGCTCTTCGGGTGCTCCAGCTCTGCCCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
Db 544 TCGGGGACGCTGNGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
Db 604 GTGGACTTTGTGCCGTAGTAGTCCATGGAAGTACTATATGCGGTCTCGG 651

## RESULT 4

## LOCUS

AR145263

DEFINITION Sequence 104 from patent US 6211338.

ACCESSION AR145263

VERSION AR145263.1 GI:15107130

## KEYWORDS

Unknown.

## SOURCE

Unclassified.

## ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolm, B.A., Taremi, S.Shane., Weber, P.C. and Yao, N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 104 03-APR-2001;

FEATURES Location/Qualifiers

source 1..1998

BASE COUNT 410 a 596 c 568 g 424 t

## ORIGIN

Alignment Scores:  
Pred. No.: 1,52e-65 Length: 1998  
Score: 881.50 Matches: 168  
Percent Similarity: 92.86% Conservatives: 14  
Best Local Similarity: 85.71% Mismatches: 11  
Query Match: 86.42% Indels: 3  
DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145263 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 64 GGTCTCTGTTATTTGTTAGTAATATTATTTATCTGCTAGTATCATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
Db 124 TCCACACAGACGGGGGCTACTTGGTTGCAAGACTAGCCTTACAGCGCGGACAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 184 AACCAAGTTCGAGGAGAGGTTTCAGTGGTTTCCACCGCAACAACTCTTCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 244 TGCCTCAACGGCGTGTGTGGACCGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
Db 304 CCAAGGGGCGCAATCACCAGATGTACACTAATGTGACCAAGACCTCGTGGCGTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 364 GCGCCCCCGGGGCGCTTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 424 GTACAGAGACATGCTGACGTCAATTCGGGTGCGCGGGCGGCGAGTAGGGGAGCGTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 484 CTCTCCCCAGGCGCTGCTCTTCTTGAAGGGCTCTTCGGGTGCTCCAGCTCTGCCCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
Db 544 TCGGGGACGCTGNGGCATCTTCCGGCTGCCGTATGCACCGGGGGGTTCGGAAGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
Db 604 GTGGACTTTGTGCCGTAGTAGTCCATGGAAGTACTATATGCGGTCTCGG 651

## RESULT 5

## LOCUS

AR145254

DEFINITION Sequence 95 from patent US 6211338.

ACCESSION AR145254

VERSION AR145254.1 GI:15107121

## KEYWORDS

Unknown.

## SOURCE

Unclassified.

AR145254 651 bp DNA linear PAT 08-AUG-2001

REFERENCE 1 (bases 1 to 651)  
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide  
 Patent: US 6211338-A 95 03-APR-2001;

JOURNAL  
 FEATURES Location/Qualifiers  
 source  
 BASE COUNT 120 a 187 c 200 g 144 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7,52e-66 Length: 651  
 Score: 878.50 Matches: 167  
 Percent Similarity: 93.33% Conservative: 15  
 Best Local Similarity: 85.64% Mismatches: 10  
 Query Match: 86.13% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x AR145254 (1-651)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTTCGTGTTATTGTTGGTAGAATATTATTTCTGTTAGTAGATACACGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 124 TCCCAACACACGCGGGCGCTACTTGGTTCGAAGAAGACTAGCTTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 ACCAGGTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACATCTCTCTGGCGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TGGCTCAACGCGGTGGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCGGCG 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 DB 304 CCAAGGGGCGCAATACCCAGATGTACACTAATGTGGACCGAGCTGCTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGCCCCCGGGCGGTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGCTCATTCGGTGGCGCGGGGGGACAGTAGGGGAGCGCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCTCCCCCAGGCTGTCTCTACTTGNAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGGCTCTTCCGGGCTGCGGTATGCACCCGGGGGTTGCGAAGCGG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAGTACTATGCGGTCTCGTCT 648

RESULT 6

AR145267  
 LOCUS AR145267 1998 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 108 from patent US 6211338.  
 ACCESSION AR145267  
 VERSION AR145267.1 GI:15107134  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)  
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide

TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide  
 Patent: US 6211338-A 108 03-APR-2001;

JOURNAL  
 FEATURES Location/Qualifiers  
 source  
 BASE COUNT 410 a 596 c 568 g 424 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,73e-65 Length: 1998  
 Score: 878.50 Matches: 167  
 Percent Similarity: 92.86% Conservative: 15  
 Best Local Similarity: 85.20% Mismatches: 11  
 Query Match: 86.13% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x AR145267 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTTCGTGTTATTGTTGGTAGAATATTATTTCTGTTAGTAGATACACGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 124 TCCCAACACACGCGGGCGCTACTTGGTTCATCAAGACTAGCTTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 ACCAGGTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACATCTCTCTGGCGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TGGCTCAACGCGGTGGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCGGCG 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 DB 304 CCAAGGGGCGCAATACCCAGATGTACACTAATGTGGACCGAGCTGCTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GCGCCCCCGGGCGGTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGCTCATTCGGTGGCGCGGGGGGACAGTAGGGGAGCGCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCTCCCCCAGGCTGTCTCTACTTGNAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGACGCTGTGGGCTCTTCCGGGCTGCGGTATGCACCCGGGGGTTGCGAAGCGG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAGTACTATGCGGTCTCGC 651

RESULT 7

AR145262  
 LOCUS AR145262 1998 bp DNA linear PAT 08-AUG-2001  
 DEFINITION Sequence 103 from patent US 6211338.  
 ACCESSION AR145262  
 VERSION AR145262.1 GI:15107129  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)  
 AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 103 03-APR-2001;

FEATURES

Location/Qualifiers

source

1..1998

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:

Pred. No.: 3.33e-65 Length: 1998  
Score: 877.50 Matches: 167  
Percent Similarity: 92.86% Conservative: 15  
Best Local Similarity: 85.20% Mismatches: 11  
Query Match: 86.03% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45262 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTCTGTGTTATTTGGTAGAATTTATTTCTGTTAGTGTAGTATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
:::|||||  
Db 124 TCCCAACAGACCGCGGCTCTAGTGGTTGCAAGATCCTAGCCTTACAGCGCGGACAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 184 AACCAGGTCGAGGAGAGGTTCCAGGTGTTCCACCGCAACACAATCCTTCTCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81  
:::|||||  
Db 244 TCGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTGCTCAAGACCTTAGCGCGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
|||||  
Db 304 CCAAGGGCCAAATCACCAGATGTACATAATGTGACGAGACCTCGTCGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 GCGCCCGCGGGCGGTCTCTTGCACCATGCACTGTGCGAGCTGCGACCTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCACGAGACATGCTGACGCTATTCGGTGGCGGGCGGCGACAGTAGGGGAGCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCTCCCCAGGCTGTCTCTACTTGAAGGCTCTGCTGGTGTCTGCTGCTGCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaValSerThrArgGlyValAlaLysAla 181  
:::|||||  
Db 544 TCGGGGACGCTGTGGCATCTTCCGGCTGCCGTATGCACCGGGGGTTGCGAAGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
|||||  
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCG 651

RESULT 8

ARI45258

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

LOCUS ARI45258 651 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 99 from patent US 6211338.  
ACCESSION ARI45258  
VERSION ARI45258.1 GI:15107125  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 651)  
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 99 03-APR-2001;  
FEATURES Location/Qualifiers

source 1..651

/organism="unknown"

BASE COUNT 120 a 187 c 200 g 144 t

ORIGIN

Alignment Scores:

Pred. No.: 1.35e-65 Length: 651  
Score: 875.50 Matches: 166  
Percent Similarity: 93.33% Conservative: 16  
Best Local Similarity: 85.13% Mismatches: 10  
Query Match: 85.83% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45258 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTCTGTGTTATTTGGTAGAATTTATTTCTGTTAGTGTAGTATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
:::|||||  
Db 124 TCCCAACAGACCGCGGCTCTAGTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 184 AACCAGGTCGAGGAGAGGTTCCAGGTGTTCCACCGCAACACAATCCTTCTCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuThrValThrValHisGlyAlaGlyThrArgThrIleAlaSer 81  
:::|||||  
Db 244 TCGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTGCTCAAGACCTTAGCGCGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
|||||  
Db 304 CCAAGGGCCAAATCACCAGATGTACATAATGTGACGAGACCTCGTCGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 GCGCCCGCGGGCGGTCTCTTGCACCATGCACTGTGCGAGCTGCGACCTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCACGAGACATGCTGACGCTATTCGGTGGCGGGCGGCGACAGTAGGGGAGCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCTCCCCAGGCTGTCTCTACTTGAAGGCTCTGCTGGTGTCTGCTGCTGCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaValSerThrArgGlyValAlaLysAla 181  
:::|||||  
Db 544 TCGGGGACGCTGTGGCATCTTCCGGCTGCCGTATGCACCGGGGGTTGCGAAGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648

RESULT 9

ARI45253

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

LOCUS ARI45253 651 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 94 from patent US 6211338.  
ACCESSION ARI45253  
VERSION ARI45253.1 GI:15107120  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 651)  
AUTHORS Malcolm,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 94 03-APR-2001;  
FEATURES Location/Qualifiers



BASE COUNT 119 a 188 c 199 g 145 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,64e-65 Length: 651  
Score: 874.50 Matches: 167  
Percent Similarity: 92.82% Conservative: 14  
Best Local Similarity: 85.64% Mismatches: 11  
Query Match: 85.74% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45253 (1-651)

```
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTCTCTGTTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCAGCGCTAC 123
|||||
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGACGCGGGGCTTACTTGGTTCATCAAGACTAGCCTTACAGCGCGGACAAG 183
|||||
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACAGGTGAGGAGAGGTTTCAAGTGGTTTCACCGCAACAAATCTCTCTGGGACC 243
|||||
Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TGGCTCAACGCGGTGCTTGGACCGTTTACCATGCTGGCTCAAGACCTTAGCGCGC 303
|||||
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCCAGATGTACACTAATGTGGACCAAGACCTCTCGCGTGGCAG 363
|||||
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGCTTCTTGACACCATGACCTGTGGCGAGCTCAGACCTTACTTG 423
|||||
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCAGAGACATGCTGACGTCATTCGGTGGCGGGCGGACAGTACAGGAGCGCTG 483
|||||
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCGAGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTGCGCT 543
|||||
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGGCATCTTCCGGGCTGCCGTATGCACCGGGGGGTTGCAAGGCG 603
|||||
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
Db 604 GTGGACTTGTGGCGGTAGAGTCCATGGAACACTACTATGCGGTCTCGTCT 648
```

RESULT 10  
ARI45266  
LOCUS ARI45266 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 107 from patent US 6211338.  
ACCESSION ARI45266  
VERSION ARI45266.1 GI:15107133  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE  
1 (bases 1 to 1998)  
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 107 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..1998  
/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t  
ORIGIN

Alignment Scores:

Pred. No.: 5.98e-65 Length: 1998  
Score: 874.50 Matches: 166  
Percent Similarity: 92.86% Conservative: 16  
Best Local Similarity: 84.69% Mismatches: 11  
Query Match: 85.74% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x ARI45266 (1-1998)

```
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
Db 64 GGTCTCTGTTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCAGCGCTAC 123
|||||
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
|||||
Db 124 TCCCAACAGACGCGGGGCTTACTTGGTTCGAAAGATCCTAGCCTTACAGCGCGGACAAG 183
|||||
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
Db 184 AACAGGTGAGGAGAGGTTTCAAGTGGTTTCCACCGCAACAAATCTCTCTGGGACC 243
|||||
Qy 62 SerIleAsnGlyValLeuThrPrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
Db 244 TGGCTCAACGCGGTGCTTGGACCGTTTACCATGCTGGCTCAAGACCTTAGCGCGC 303
|||||
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
|||||
Db 304 CCAAGGGGCAATCCAGATGTACACTAATGTGGACCAAGACCTCTCGCGTGGCAG 363
|||||
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
Db 364 GCGCCCCCGGGCGGCTTCTTGACACCATGACCTGTGGCGAGCTCAGACCTTACTTG 423
|||||
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
Db 424 GTCAGAGACATGCTGACGTCATTCGGTGGCGGGCGGACAGTACAGGAGCGCTG 483
|||||
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
Db 484 CTCTCCCGAGCGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTGCGCT 543
|||||
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
Db 544 TCGGGGACGCTGTGGGCATCTTCCGGGCTGCCGTATGCACCGGGGGGTTGCAAGGCG 603
|||||
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
Db 604 GTGGACTTGTGGCGGTAGAGTCCATGGAACACTACTATGCGGTCTCGC 651
```

RESULT 11  
ARI45261  
LOCUS ARI45261 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 102 from patent US 6211338.  
ACCESSION ARI45261  
VERSION ARI45261.1 GI:15107128  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE  
1 (bases 1 to 1998)  
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 102 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..1998  
/organism="unknown"

BASE COUNT 409 a 597 c 567 g 425 t  
ORIGIN

Alignment Scores:

Pred. No.: 7.27e-65 Length: 1998  
 Score: 873.50 Matches: 167  
 Percent Similarity: 92.35% Conservatives: 14  
 Best Local Similarity: 92.35% Mismatches: 12  
 Query Match: 85.20% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GTTCTGTGTATGTTGGTAGAATATTATCTGCTAGTGTAGTATCAGCGCTAC 123  
 |||||  
 QY 22 AlaGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 :|||  
 Db 124 TCCCAACAGACGCGGGGCTACTTGGTGTGCATCATCTAGCTTACAGGCGGACAAG 183  
 |||||  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 AACCAAGTCGAGGAGAGGTTGAGTGGTGTTCACCGCAACACAATCTCTCTGGCGACC 243  
 |||||  
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 :|||  
 Db 244 TCGCTCAACGGCGTGTGGACCGTTTACCATGGTGTGCTCAAGACCTTAGCGCGC 303  
 |||||  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 :|||  
 Db 304 CCAAGGGGCAATCACCACATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 363  
 |||||  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 364 GCGCCCCCGGGCGGTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
 |||||  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 424 GTCACGACATGCTGACGTCTATTCGGTGTGGCGGGGGGAGGAGCGCTG 483  
 |||||  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 :|||  
 Db 484 CTCTCCCCAGGCGTCTCTCTACTTGAAGGGCTCTTCGGTGTGCTCCTCTGCGCT 543  
 |||||  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 :|||  
 Db 544 TCGGGGACGCTGGGGCATCTTCGGGCTGCCGTATGCACCGGGGGGTTCGAGGCG 603  
 |||||  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 :|||  
 Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651  
 |||||

RESULT 12  
 AF268278 AR145269  
 LOCUS Sequence 110 from patent US 6211338. 2016 bp DNA linear PAT 08-AUG-2001

DEFINITION  
 ACCESSION  
 VERSION  
 AR145269.1 GI:15107136

KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unclassified.

REFERENCE  
 1 (bases 1 to 2016)  
 AUTHORS Malcolin,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide  
 JOURNAL Patent: US 6211338-A 110 03-APR-2001;  
 FEATURES Location/Qualifiers

source  
 1..2016  
 /organism="unknown"

BASE COUNT 412 a 603 c 570 g 431 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7.34e-65 Length: 2016  
 Score: 873.50 Matches: 167

Percent Similarity: 92.35% Conservatives: 14  
 Best Local Similarity: 85.20% Mismatches: 12  
 Query Match: 85.64% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-20 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgGlyLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 82 GGTCTGTGTATTGTTGGTAGAATATTATCTGCTAGTGTAGTATCAGCGCTAC 141  
 |||||  
 QY 22 AlaGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 :|||  
 Db 142 TCCCAACAGACGCGGGGCTACTTGGTGTGCATCATCTAGCTTACAGGCGGACAAG 201  
 |||||  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 202 AACCAAGTCGAGGAGAGGTTGAGTGGTGTTCACCGCAACACAATCTCTCTGGCGACC 261  
 |||||  
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 :|||  
 Db 262 TCGCTCAACGGCGTGTGGACCGTTTACCATGGTGTGGCTCAAGACCTTAGCGCGC 321  
 |||||  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 :|||  
 Db 322 CCAAGGGGCAATCACCACATGTACACTAATGTGGACAGGACCTCGTGGCTGGCAG 381  
 |||||  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 382 GCGCCCCCGGGCGGTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 441  
 |||||  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 442 GTCACGACATGCTGACGTCTATTCGGTGTGGCGGGGGGAGGAGCGCTG 501  
 |||||  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 :|||  
 Db 502 CTCTCCCCAGGCGTCTCTCTACTTGAAGGGCTCTTCGGTGTGCTCCTCTGCGCT 561  
 |||||  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 :|||  
 Db 562 TCGGGGACGCTGGGCACTCTCCGGGCTGCCGTATGCACCGGGGGGTTCGAGGCG 621  
 |||||  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 :|||  
 Db 622 GTGGACTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 669  
 |||||

RESULT 13  
 AF268278

LOCUS AF268278 12734 bp RNA linear VRL 12-JUL-2000

DEFINITION Pestivirus type 1, complete genome.  
 ACCESSION AF268278

VERSION AF268278.1 GI:9049956  
 KEYWORDS  
 SOURCE

ORGANISM  
 Pestivirus type 1  
 Pestivirus type 1

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Pestivirus.

REFERENCE  
 1 (bases 1 to 12734)

AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassilev,V.,  
 Denis,R.O., Hong,Z. and Lau,J.Y.

TITLE Generation and characterization of a hepatitis C virus NS3  
 protease-dependent bovine viral diarrhea virus

JOURNAL J. Virol. 74 (14), 6339-6347 (2000)  
 MEDLINE 20323484  
 PUBMED 10864644

REFERENCE  
 2 (bases 1 to 12734)

AUTHORS Lai,V.C.B. and Hong,Z.  
 TITLE Direct Submission

JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research  
 Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

FEATURES  
 Location/Qualifiers  
 1..12734  
 source



protease and NS4A cofactor peptide  
Patent: US 6211338-A 98 03-APR-2001;

JOURNAL

FEATURES

Source 1. .651  
Location/Qualifiers  
BASE COUNT 119 a 188 c 199 g 145 t  
ORIGIN /organism="unknown"

Alignment Scores:

Pred. No.: 2,95e-65 Length: 651  
Score: 871.50 Matches: 166  
Percent Similarity: 92.82% Conservative: 15  
Best Local Similarity: 85.13% Mismatches: 11  
Query Match: 85.44% Indels: 3  
DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145257 (1-651)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTTCGTGTTATTCTTGGTAGAATTAATTTATCTGGTAGTATCATCGGCTTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
|||||
DB 124 TCCCAACAGACGGCGGCTTACTTGGTCATCAAGACTAGCTTACAGCGCGGACAAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 AACAGGTCCAGGAGAGGTTTCAGGTGGTTCCACGGCAACACAAATCCTTCCTGGCGACC 243
QY 62 SerIleAsnGlyValLeuThrValThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TGGGTCAACGGCGTGTGTTGGACCGTTTACCATGGTGTGGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
DB 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGGACCTCGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 364 CGCGCCCGCGGGCGGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 424 GTCAGGACATGCTGACGTCATTCGGGTGCGCGCGGGCGGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCysPro 161
|||||
DB 484 CTCCTCCCGAGGCTGTCTCTACTTCAAGGGCTCTGCTGGTGTCTGCTGCTGCTCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
DB 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGTTCGAAGGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
DB 604 GTGGACTTTGTGCCGTAGACTCCATGGAAACTACTATCGCGTCT 648
```

RESULT 15

AR145252

LOCUS

DEFINITION

AR145252

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AR145252 651 bp DNA linear PAT 08-AUG-2001

Sequence 93 from patent US 6211338.

AR145252.1 GI:15107119

Unknown.

Unclassified.

1 (bases 1 to 651)

Malcolm, B.A., Taremi, S., Shane, J., Weber, P.C. and Yao, N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 93 03-APR-2001;

FEATURES  
Source 1. .651  
Location/Qualifiers

BASE COUNT 119 a 188 c 199 g 145 t  
ORIGIN /organism="unknown"

Alignment Scores:

Pred. No.: 3.59e-65 Length: 651  
Score: 870.50 Matches: 166  
Percent Similarity: 92.82% Conservative: 15  
Best Local Similarity: 85.13% Mismatches: 11  
Query Match: 85.34% Indels: 3  
DB: Gaps: 1

US-09-965-594-20 (1-197) x AR145252 (1-651)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTTCGTGTTATTCTTGGTAGAATTAATTTATCTGGTAGTATCATCGGCTTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
|||||
DB 124 TCCCAACAGACGGCGGCTTACTTGGTTCACAGATCACTAGCCTTACAGCGCGGACAAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 AACAGGTCCAGGAGAGGTTTCAGGTGGTTCCACGGCAACACAAATCCTTCCTGGCGACC 243
QY 62 SerIleAsnGlyValLeuThrValThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TGGGTCAACGGCGTGTGTTGGACCGTTTACCATGGTGTGGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
DB 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGGACCTCGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 364 CGCGCCCGCGGGCGGTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 424 GTCAGGACATGCTGACGTCATTCGGGTGCGCGCGGGCGGACAGTAGGGGAGCCTG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCysPro 161
|||||
DB 484 CTCCTCCCGAGGCTGTCTCTACTTCAAGGGCTCTTCCGGGTGGTCCACTGCTGCTCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
DB 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTATGCACCGGGGGTTCGAAGGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
DB 604 GTGGACTTTGTGCCGTAGACTCCATGGAAACTACTATCGCGTCT 648
```

Search completed: August 31, 2003, 00:46:32  
Job time : 2569.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-20

Perfect score: 1020

Sequence: 1 MKKKGWVIVGRINLSGDFA.....VAKAVDFIPVESLETMRSP 197

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=slp  
-O=/cnp2\_1/USPTO.spool/US09965594/runat\_29082003\_151918\_28302/app\_query.fasta\_1.2872  
-DB=N\_Geneseq\_19Jun03 -CFMT=fastep -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cd1  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN\_1\_1\_1412 /runat\_29082003\_151918\_28302 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_19Jun03:\*\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1020	100.0	594	21	AAA73333	Hepatitis C virus
2	1010	99.0	594	21	AAA73334	Hepatitis C virus
3	1005	98.5	594	21	AAA73332	Hepatitis C virus
4	990	97.1	594	21	AAA73331	Hepatitis C virus
5	973	95.4	594	21	AAA73330	Hepatitis C virus
6	946	92.7	594	21	AAA73335	Hepatitis C virus
7	939	92.1	588	21	AAA73329	Hepatitis C virus
8	912	89.4	588	21	AAA73328	Hepatitis C virus
9	892.5	87.5	12734	24	ABA95615	Chimeric BVDV/HCV
10	885.5	86.8	1998	20	AAH80355	HCV NS4A-NS3 compl
11	882.5	86.5	1998	20	AAH80359	HCV NS4A-NS3 compl
12	881.5	86.4	1998	20	AAH80354	HCV NS4A-NS3 compl
13	878.5	86.1	651	20	AAH80345	HCV NS4A-NS3 compl
14	878.5	86.1	1998	20	AAH80358	HCV NS4A-NS3 compl
15	877.5	86.0	1998	20	AAH80353	HCV NS4A-NS3 compl
16	875.5	85.8	651	20	AAH80349	HCV NS4A-NS3 compl
17	874.5	85.7	612	25	ABX15706	Anti-viral synthet
18	874.5	85.7	651	20	AAH80344	HCV NS4A-NS3 compl
19	874.5	85.7	1998	20	AAH80357	HCV NS4A-NS3 compl
20	873.5	85.6	1998	20	AAH80352	HCV NS4A-NS3 compl
21	873.5	85.6	2013	20	AAH80360	HCV NS4A-NS3 compl
22	871.5	85.4	651	20	AAH80348	HCV NS4A-NS3 compl
23	870.5	85.3	651	20	AAH80343	HCV NS4A-NS3 compl
24	870.5	85.3	1998	20	AAH80356	HCV NS4A-NS3 compl
25	870.5	85.3	2016	20	AAH80361	HCV NS4A-NS3 compl
26	870	85.3	648	20	AAH80365	HCV NS4A-NS3 compl
27	868	85.1	648	20	AAH80363	HCV NS4A-NS3 compl
28	867.5	85.0	650	20	AAH80347	HCV NS4A-NS3 compl
29	867.5	85.0	651	20	AAH80351	HCV NS4A-NS3 compl
30	866.5	85.0	651	20	AAH80342	HCV NS4A-NS3 compl
31	864	84.7	648	20	AAH80362	HCV NS4A-NS3 compl
32	863.5	84.7	650	20	AAH80346	HCV NS4A-NS3 compl
33	863.5	84.7	651	20	AAH80350	HCV NS4A-NS3 compl
34	861	84.4	8145	20	AAH23259	Plasmid PET-BS(+)/
35	859	84.2	1933	20	AAH23258	HCV NS3 DNA. Hepa
36	858.5	84.2	9646	19	AAV59361	Hepatitis C virus
37	858.5	84.2	9646	24	ABK87285	Hepatitis C virus
38	858.5	84.2	12980	19	AAV59364	Hepatitis C virus
39	858.5	84.2	12980	24	ABK87286	Hepatitis C virus
40	858.5	84.2	16622	21	AAZ36212	Nucleotide sequenc
41	854.5	83.8	5300	10	AAH92097	Combined open read
42	854.5	83.8	5360	10	AAH90327	Hepatitis C virus
43	854.5	83.8	6905	10	AAH92103	Combined open read
44	854.5	83.8	7310	10	AAH92106	Combined open read
45	854.5	83.8	7310	10	AAH90336	Composite hepatiti

ALIGNMENTS

RESULT 1  
AAA73333  
ID AAA73333 standard; DNA; 594 BP.  
XX AAA73333;  
AC AAA73333;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #6.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein; ds.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers



XX  
SQ Sequence 594 BP; 105 A; 192 C; 151 G; 146 T; 0 other;

Alignment Scores:	6.24e-85	Length:	594
Pred. No.:	Score:	Matches:	196
	1010.00	Conservative:	0
Percent Similarity:	99.49%	Mismatches:	1
Best Local Similarity:	99.49%	Indels:	0
Query Match:	99.02%	Gaps:	0
DB:	21		

US-09-965-594-20 (1-197) x AAA73334 (1-594)

Qy	1	MetLysLysLysGlySerValValleValGlyArgIleAsnLeuSerGlyAspThrAla	20
Db	1	ATGAATAAAAAAGATCGTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGT	60
Qy	21	TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyAcgAsp	40
Db	61	TACGCTCAGCAGACTCGAGGTGACGAGGTACCCAGAGACCTCCCAACCCGGTCTGTGAC	120
Qy	41	LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla	60
Db	121	AAAAACACAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTACCCAGACCTTCCTGGCT	180
Qy	61	ThrSerIleAsnGlyValLeuTrpThrValTrpHisGlyAlaGlyThrArgThrIleAla	80
Db	181	ACCTCCATCAACGGTGTTCTGTGGACCGTTTACCACACGGTGTGGTACCGGTACCATCGT	240
Qy	81	SerProLysGlyProValThrGlnMetTrpThrAsnValAspLysAspLeuValGlyTrp	100
Db	241	TCCCGGAAAGGTCGGTTACCCAGATGTACCAACGTTGACAAAGACCTGGTTGGTTGG	300
Qy	101	GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTrp	120
Db	301	CAGGCTCCGACGGTTCCTCCGTACCCCGCGACCTGCGGTCTCTCCGACCTGTAC	360
Qy	121	LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer	140
Db	361	CTGGTTTACCCGTCAGCGTGAGCTTATCCCGGTTTCGTGCTGTGAGCTCCCGTGGTTC	420
Qy	141	LeuLeuSerProArgProIleSerTyrIleuLysGlySerSerGlyGlyProLeuLeuCys	160
Db	421	CTGCTGTCTCCCGCGTCCGATCTCTTACCTGAAGAGGTTCTCCCGGTGGTCCGCTGTGC	480
Qy	161	ProAlaGlyHisAlaValGlyIlePheArqAlaAlaValSerThrArgGlyValAlaLys	180
Db	481	CCGGCTGGTCAGCCTGTGGTATCTTCCTGCTGCTGTTTCCACCCGCTGGTGTGCTATA	540
Qy	181	AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	197
Db	541	GTGTGTGATCTCATCCCGGTTGAATCTCTCGTGAAGAACCACTGCGTTCCTCCGG	591

## RESULT 3

AAA73332

ID AAA73332 standard; DNA; 594 BP.

XX

AC AAA73332;

XX

DT 19-DEC-2000 (first entry)

**CXX**

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.

XX  
KW Hepatitis: NS3 protease: viral replication: chronic liver disease:

**KW** liver failure:  
**NM** hepatitis;  
process; viral  
reproduction;  
cancer; mutant;

[illegible]

OS Hepatitis C virus.

OS Synthetic.

XX

FH	Key	Location/Qualifiers
FM		

FT	CDS
1.1594	
1.1594	

```

Fi      /tag= a
Ft      /product= "NS4A-NS3 fusion protein" $S

```

XX	WO200040707-A1.
XX	13-JUL-2000.
XX	
XX	06-JAN-2000; 2000WO-US00345.
XX	
XX	08-JAN-1999; 99US-0115271.
XX	
XX	(BRIM ) BRISTOL-MYERS SQUIBB CO.
XX	
XX	Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX	WPI: 2000-465976/40.
XX	P-PSDB; AAB15223.
XX	
XX	Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
PT	substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
PT	amino acid, useful for screening inhibitors that may treat hepatitis C
PT	.
XX	
XX	Claim 26; Fig 15; 66pp; English.
XX	
XX	The present sequence is the coding sequence for a mutated version of a
CC	fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
CC	protease enzymes. These proteins are both essential for the replication
CC	of the virus, acting to cleave its replicative proteins from the
CC	polyprotein produced from the HCV genome. Inhibitors of the two proteins
CC	should be effective as antiviral treatments of HCV infection. This is
CC	useful as HCV can lead to chronic liver disease such as cirrhosis, liver
CC	failure and liver cancer. The present invention concerns a number of NS3
CC	mutants and NS3-NS4A fusion proteins which can be used to identify
CC	inhibitors of this type, as well as enabling structural studies of the
CC	protease and protease-inhibitor complexes. The protein produced from this
CC	sequence contains the alpha-helix0-1 variant.
XX	
XX	Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:

Pred. No.:	1,84e-85	Length:	594
Score:	1005.00	Matches:	194
Percent Similarity:	99.49%	Conservative:	2
Best Local Similarity:	98.48%	Mismatches:	1
Query Match:	98.53%	Indels:	0
DB:	21	Gaps:	0

US-09-965-594-20 (1-197) x AAA73332 (1-594)

[illegible]



```

Db 361 CTGGTTACCGTCACGTCACGTTATCCGGTTCGTGCTCGTGGTACCTCCGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160
Db 421 CTGCTGTCCCGCGCCGATCTCTACCTGAAGGTTCTCCGGTGGTCCGCTGCTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTGGTATCTTCGCTGCTGCTGCTTCACCGGTGGTGTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCGCCG 591

RESULT 4
AAA73331
ID AAA73331 standard; DNA: 594 BP.
AC AAA73331;
DT 19-DEC-2000 (first entry)
XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #4.
DE Hepatitis C virus NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutein; ds.
XX Hepatitis C virus.
OS Synthetic.
XX Key Location/Qualifiers
FH 1.594
FT CDS /*tag= a
FT /product= "NS4A-NS3 fusion protein #4"
XX WO200040707-A1.
XX 13-JUL-2000.
XX 06-JAN-2000: 2000WO-US00345.
XX 08-JAN-1999: 99US-0115271.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
XX WPI: 2000-465976/40.
XX P-PSDB; AAB15222.
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1
XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic
XX amino acid, useful for screening inhibitors that may treat hepatitis C
XX
XX Claim 26: Fig 14; 66pp; English.
XX
XX The present sequence is the coding sequence for a mutated version of a
XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A
XX protease enzymes. These proteins are both essential for the replication
XX of the virus, acting to cleave its replicative proteins from the
XX polypeptide produced from the HCV genome. Inhibitors of the two proteins
XX should be effective as antiviral treatments of HCV infection. This is
XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver
XX failure and liver cancer. The present invention concerns a number of NS3
XX mutants and NS3-NS4A fusion proteins which can be used to identify
XX inhibitors of this type, as well as enabling structural studies of the
XX protease and protease-inhibitor complexes. The protein produced from this
XX sequence contains the alpha-helix0-1 variant.
XX
XX Sequence 594 BP; 105 A; 187 C; 155 G; 147 T; 0 other;

```

```

Alignment Scores: 4.72e-84 Length: 594
Pred. No.: 990.00 Matches: 191
Score: 990.00
Percent Similarity: 97.97% Conservatives: 2
Best Local Similarity: 96.95% Mismatches: 4
Query Match: 97.06% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-20 (1-197) x AAA73331 (1-594)
QY 1 MetLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
Db 1 ATGAAAAAAGGATCCGTTGTTATCGTCGCCGCTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAsp 40
Db 61 TAGCTCAGCAGACTCGAGGTGAGGAGGTGCGAAGAAACCTCCAGACCGCTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
Db 121 AAAAACGAGGTTCAGGTGAAGTTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGCT 180
QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
Db 181 ACCTGATCAACGGTGTTCGTCGACCGCTTACCCAGCGTCTGGTACCGTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
Db 241 TCCCGAAGGTCGCGTTACCCAGATGTACACCAAGTTGACAAAGACCTGCTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
Db 301 CAGGCTCCGCGAGGTTCCCGTTCCTCCGACCGCTGACCTGGCGTTCCTCCAGCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
Db 361 CTGTTACCGCTCAGCGTGACGTTATCCCGTTCGTCGTCGTCGTCGTCGTCGTCGTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160
Db 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCTCCGGTGGTCCGCTGCTGTC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
Db 481 CCGGCTGGTCACGCTGTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 541 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCGCCG 591

RESULT 5
AAA73330
ID AAA73330 standard; DNA: 594 BP.
XX AC AAA73330;
XX DT 19-DEC-2000 (first entry)
XX DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #3.
XX KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
XX liver failure; liver cancer; mutant; mutein; ds.
XX OS Hepatitis C virus.
XX Synthetic.
XX Key Location/Qualifiers
FH 1.594
FT CDS /*tag= a
FT /product= "NS4A-NS3 fusion protein #3"
XX WO200040707-A1.

```



PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15221.  
 XX  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX Claim 26; Fig 13; 66pp; English.  
 PS  
 XX The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0-1 variant.  
 XX  
 XX Sequence 594 BP; 103 A; 186 C; 156 G; 149 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,87e-82 Length: 594  
 Score: 973.00 Matches: 188  
 Percent Similarity: 96.45% Conservative: 2  
 Best Local Similarity: 95.43% Mismatches: 7  
 Query Match: 95.39% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-20 (1-197) x AAA73330 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGAGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCCGGTGACCGCT 60  
 QY 21 TyrAlaGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAsp 40  
 DB 61 TAGCGCTCAGCAGACTCGAGGTGAGGAGGGTGCACAAAGAACCTCCAGACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAAACCAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 180  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValThrValHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTGCATCAACGCTGTTGCTGGACCGTTTACACGGGTGCTGTTACCGGTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTyr 100  
 DB 241 TCCCGGAAGAGTCCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGGTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CCGGCTCCGAGGGTCCCGTTCCTGACCCCGGTGACCTGCGGTTCCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGTTTACCGGTGACGTGACGTATATCCCGGTTTCGTCGTCGTCGTCGTCGTCGTCGTC 420

QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGGTCGGATCTCTACTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 481 CCGGCTGGTCACGCTGTGGTATCTTCCGTCGTGCTGTTTGCACCCGCTGCTGCTAA 540  
 QY 181 AlavalAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCCTGTGACTTCATCCCGGTGTAATCCCTGGAAACCAACCATGCTGCTCCCG 591

RESULT 6  
 AAA73335  
 ID AAA73335 standard; DNA: 594 BP.  
 XX  
 AC AAA73335;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #8.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutain; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT CDS 1..594  
 FT /\*tag= a  
 FT /product= "NS4A-NS3 fusion protein #8"  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15226.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Disclosure; Fig 18; 66pp; English.  
 XX  
 XX The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0 wild-type sequence.  
 XX  
 SQ Sequence 594 BP; 98 A; 189 C; 153 G; 154 T; 0 other;

Alignment Scores:  
 Pred. No.: 6.41e-80 Length: 594  
 Score: 946.00 Matches: 186

Percent Similarity: 94.42% Conservative: 0  
 Best Local Similarity: 94.42% Mismatches: 11  
 Query Match: 92.75% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-20 (1-197) x AAA73329 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGAGTCCGTTGTTATCGTCGCCGATCAACCTGTCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAsp 40  
 DB 61 TACGCTCAGCAGACTCGAGGTCGTGGTGTGCATCATCACCTCCCTGACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAAACCCAGGTGAAGGTGAAGTTCAGATGTTTCCACCGCTGCTCAGACCTTCCTGGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTGCATCAACGGGTGTTGCTGGACCGTTTACACAGGTGCTGTACCGGTACCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
 DB 241 TCCCGGAAGGTCCGGTTATCCAGATGTACACACAGTGTGACAAGACCTGGTGGTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CGCGCTCCGAGGTTCCCGTCCCTGACCGCTGCACCTCGCTTCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGTTATCCCGTACGCTGACGTTATCCCGGTTCGTCGTGGTGACTCCCGGTGCTC 420  
 QY 141 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuLys 160  
 DB 421 CTGCTGTCGCCCGCTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGC 480  
 QY 161 ProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 481 CGCGCTGGTACGCTGTTGGTATCTTCGCTGCTGCTGTTGACCCCGGTGTTGCTAAA 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCTGTTGACTTATCCCGGTTGAATCTCCCTGGAACACCATGCGTTCCCG 591

## RESULT 7

AAA73329

ID AAA73329 standard; DNA; 588 BP.

XX AC AAA73329;

AC AAA73329;

XX 19-DEC-2000 (first entry)

DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;

KW liver failure; liver cancer; mutant; mutein; ds.

XX Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

PH 1..588

FT /tag= a

FT /product= "NS4A-NS3 fusion protease protein #2"

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 XX P-PSDB; AAB15220.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 XX substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 XX amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX  
 XX Claim 26; Fig 12; 66pp; English.  
 XX The present sequence is the coding sequence for a mutated version of a  
 XX fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 XX protease enzymes. These proteins are both essential for the replication  
 XX of the virus, acting to cleave its replicative proteins from the  
 XX polyprotein produced from the HCV genome. Inhibitors of the two proteins  
 XX should be effective as antiviral treatments of HCV infection. This is  
 XX useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 XX failure and liver cancer. The present invention concerns a number of NS3  
 XX mutants and NS3-NS4A fusion proteins which can be used to identify  
 XX inhibitors of this type, as well as enabling structural studies of the  
 XX protease and protease-inhibitor complexes. The protein produced from this  
 XX sequence contains the alpha-helix0-1 variant.  
 XX Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

## Alignment Scores:

Pred. No.: 2,88e-79 Length: 588  
 Score: 939.00 Matches: 184  
 Percent Similarity: 94.92% Conservative: 3  
 Best Local Similarity: 93.40% Mismatches: 8  
 Query Match: 92.06% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-20 (1-197) x AAA73329 (1-588)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGAGTCCGTTGTTATCGTCGCCGATCAACCTGTCGGTGACACCGCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAsp 40  
 DB 55 TACGCTCAGCAGACTCGAGTGGAGGTTGCCAAGAAACCTCCAGACCGGTGCTGAC 114  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 115 AAAAACCCAGGTGAAGGTGAAGTTCAGATGTTTCCACCGCTGCTCAGACCTTCCTGGCT 174  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 175 ACCTGCATCAACGGGTGTTGCTGGACCGTTTACACAGGTGCTGTACCGGTACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
 DB 235 TCCCGGAAGGTCCGGTTATCCAGATGTACACCAAGTGTGACAAGACCTGTTGTTGG 294  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 295 CCGGCTCCGAGGTTCCCGTTCCTGACCCCGTGCACCTGCGGTTCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 355 CTGCTTACCGCTCAGCTACGCTTATCCCGGTTCGTCGTCGTGGTGACTCCCGTGGTTC 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyProLeuLeuLys 160  
 DB 415 CTGCTGTCGCCCGCTCCGATCTCTACCTGAAAGGTTCCTCCCGTGGTCCGCTGCTGTC 474

QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 475 CCGCGTGGTCACCGCTGTGTGTATCTTCGCGTGTGTGTTGCACCGGTGTGTGCTATAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAACACCATGCTGCCCG 585

RESULT 8  
 AAA73328  
 ID AAA73328 standard; DNA: 588 BP.  
 XX  
 AC AAA73328:  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..588 /\*tag= a  
 FT /\*product= "NS3-NS4A fusion protein"  
 XX  
 PN W0200040707-Al.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA  
 XX Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 PI  
 XX WPI: 2000-465976/40.  
 DR P-PSDB: AAB15212.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Disclosure; Fig 10; 66pp; English.  
 XX  
 CC The present sequence is the coding sequence for a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polypeptide produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes.  
 XX  
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.89e-77 Length: 588  
 Score: 912.00 Matches: 182  
 Percent Similarity: 92.89% Conservative: 1  
 Best Local Similarity: 92.39% Mismatches: 12  
 Query Match: 89.41% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-20 (1-197) x AAA73328 (1-588)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGGTTCCGTTGTTATCTCGCGCGGTATAGTACTGACCGGT-----GCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnLysThrSerHisThrSerHisThrGlyArgAsp 40  
 DB 55 TACGCTCAGCAGACTCGAGGTCGTGGTTCATCATCCTCCCTGACCGGTGCTGAC 114  
 QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 115 AAAAACCAGGTTGAAGTGAAGTTCAGATCGTTCCACCGCTGCTCAGACCTCTCTGCT 174  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 175 ACTGCTCATCAGCGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCGGTACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
 DB 235 TCCCGAAGGTCGCGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTTGGTTGG 294  
 QY 101 GlnAlaProGlnGlnSerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 295 CCGGCTCCGACGGTTCGCGTTCCCTGACCGCGTGCACGTGCGGTTCCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 355 CTGGTTACCGGTCACGCTGACGTTATCCCGGTTCGTCGCTGGTGTGCTCCGTTGCC 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 415 CTGCTGTCCCGCGTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGTC 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 475 CCGGCTGGTTCACGCTGTTGTTATCTTCGCTGCTGCTGTTGTCACCGGTGTGCTAAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCATCCCGGTGTAATCCCTGGAACACCATGCTGCCCG 585

RESULT 9  
 ABA95615  
 ID ABA95615 standard; DNA: 12734 BP.  
 XX  
 AC ABA95615;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Chimeric BVDV/HCV NS3-wt sequence.  
 XX  
 KW Pestivirus; Npro; protease; NS3; screening; ds.  
 XX  
 OS Chimeric - Bovine viral diarrhea virus.  
 OS Chimeric - Hepatitis C virus.  
 XX  
 PN US6326137-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 PF 25-JUN-1999; 99US-0344456.  
 XX  
 PR 25-JUN-1999; 99US-0344456.  
 XX  
 XX (SCHE ) SCHERING CORP.  
 PA  
 XX Hong Z, Lai VCH, Lau JYN;  
 PI WPI: 2002-121103/16.  
 DR  
 XX Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)  
 PT

PT pestivirus genome where the Npro protease gene is replaced with NS3  
 PT protease gene, useful for in vivo screening of compounds which inhibit  
 PT HCV infection

XX Example 2: Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a  
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
 CC each junction site recognised by the Npro protease is replaced with a  
 CC junction site recognised by the HCV NS3 protease. The construct is useful  
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
 CC protease, where screening may be in cell culture or in an animal model.  
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.12e-73 Length: 12734  
 Score: 892.50 Matches: 177  
 Percent Similarity: 92.82% Conservative: 4  
 Best Local Similarity: 90.77% Mismatches: 11  
 Query Match: 87.50% Indels: 3  
 DB: 24 Gaps: 1

US-09-965-594-20 (1-197) x ABA95615 (1-12734)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 Db 413 GGTAGTGTGTTATTCTGTTAGAAATGTTTATCTGTAGTGGTAGTATCACGGCGTAC 472

Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 Db 473 GCCACGACGACGAGAGCGCTCTAGGGTGTAAAGTACACAGCTCAGCTGGCGGACAAA 532

Qy 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 Db 533 AACCAAGTGGAGGGTGGTGCAGATCGTGTCACTGTCTACCAACCTTCTGTGCNAAG 592

Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 Db 593 TGCATCAATGGGTATGTCTGAGTGTCTACACGGCGCGGAAACGAGCATCCATCA 652

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 Db 653 CCCAAGGGTCTGTCTATCCAGATGTATACCAATGTGGACCAAGAGCTTGTGGGCTGGCC 712

Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 Db 713 GTCCTCTAAGTTCGCGCTCATGTACACCTCGACCTGCGGCTCTCGGACCTTACCTG 772

Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 Db 773 GTTACGAGGACGCCGACGCTATTCCCGTCCGCGCGAGGTGATAGCAGGGGTAGCCCTG 832

Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 Db 833 CTTTCG 892

Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaIleValSerThrArgGlyValAlaLysAla 181  
 Db 893 GCGGGACACGCCGCGCGCTATTACGCGCGCGCGGTGTGCACCGGTGGAGTGGCCAGGCG 952

Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 Db 953 GTGGACTTTATCCCTGTGGAGAACCTTAGAGACAACCATGAGATCC 997

RESULT 10

AAx80355

ID AAx80355 standard; cDNA; 1998 BP.

XX

AAx80355;

07-SEP-1999 (first entry)

HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.

HCV; hepatitis C virus; single chain recombinant complex; linker;  
 NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 hydrophobic domain; covalent complex; detection; inhibitor; ss.

Hepatitis C virus.

Synthetic.

W09928482-A2.

10-JUN-1999.

24-NOV-1998; 98WO-US24528.

28-JUL-1998; 98US-0094331.

28-NOV-1997; 97US-0067315.

(SCHE ) SCHERING CORP.

Malcolm BA, Taremi SS, Weber PC, Yao N;

WPI; 1999-385385/32.

New hepatitis C virus covalent complexes

Disclosure; Page 166-169; 21pp; English.

The present invention describes a covalent hepatitis C virus (HCV)  
 NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 to the amino terminus of the HCV NS3 protease domain. The present  
 sequence encodes an example of the above complex. The covalent  
 NS4A-NS3 complexes are useful for structural determination and  
 determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 They can also be used for detecting inhibitors of the protease activity.  
 the helicase activity and the ATPase activity of NS3. The covalent  
 NS4A-NS3 complexes are more soluble, stable and active than the non-  
 covalent protease-peptide complexes previously available.

Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:

Pred. No.: 1.41e-73 Length: 1998  
 Score: 885.50 Matches: 168  
 Percent Similarity: 93.37% Conservative: 15  
 Best Local Similarity: 85.71% Mismatches: 10  
 Query Match: 86.81% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAx80355 (1-1998)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 Db 64 GGTTCTGTCTATTATTTGTGTAGAAATTTATTTATCTGTAGTATCATCGGCTAC 123

Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 Db 124 TCCCAACAGACGGGGGCGCTACTTGGTTGCAAGAACTAGCCTTACGGCCGGGCAAG 183

Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 Db 184 AACCAAGTTCGAGGAGAGGTTCAAGTGTTCACCGCAACAATCTTCTCTGGCGACC 243

Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 Db 244 TGCCTCAACGGCGTGTGTGGACCGTTTACATGTGCTGCCTCAAGACCTTAGCGCGC 303

Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGACCTCTCGGCTGGCAG 363  
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 Db 364 GCGCCCCCGGGGGGGTCTCTTGACACCATGACCTGTGGCAGCTCAGACCTTTACTTG 423  
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 Db 424 GTCACGAGACGCTGACGTTCATTCGGTCCGGCGGGGGGACAGTAGGGGAGCCCTG 483  
 Qy 142 LeuSerProArgProLysGlySerLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 Db 484 CTCCTCCCGAGGCTGCTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543  
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGGCTGCGGTATGCACCGGGGGGTTCGGAAGCG 603  
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACACTATGCGGTCTCCG 651

RESULT 11  
 AAX80359  
 ID AAX80359 standard; cDNA; 1998 BP.  
 XX AC AAX80359;  
 XX DT 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN W09928482-A2.  
 XX PD 10-JUN-1999.  
 XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.  
 XX PA (SCHE ) SCHERING CORP.  
 XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX DR WPI; 1999-385385/32.  
 XX PT New hepatitis C virus covalent complexes  
 XX PS Disclosure; Page 179-182; 21pp; English.  
 XX CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 2,69e-73 Length: 1998  
 Score: 882.50 Matches: 167  
 Percent Similarity: 93.37% Conservative: 16  
 Best Local Similarity: 85.20% Mismatches: 10  
 Query Match: 86.52% Indels: 3  
 DB: 20 Gaps: 1  
 OS-09-965-594-20 (1-197) x AAX80359 (1-1998)  
 Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 Db 64 GTTCTGTGTTATTGTTGTTAGAAATTATTTATCTGTAGTGTAGTATCAGGCGCTAC 123  
 Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 Db 124 TCCCAACAGACGCGGGGCTACTTGTGTTGCAAGAGACTAGCCTTACAGCGCGGACAG 183  
 Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 Db 184 AACCAAGTTCAGGAGAGGTTTCAGTGGTTCCACCGCACACAAATCCTTCCTGGCGACC 243  
 Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 Db 244 TCGCTCAACGCGCTGTGTGGACGCTTACCATGCTGTGCTCAAGACCTTAGCGCGC 303  
 Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACAGACCTCTCGGCTGGCAG 363  
 Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 Db 364 GCGCCCCCGGGGGCGTCTCTGACACATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 Db 424 GTCACGAGACATGCTGACGTTCATTCGGTCCGGCGGGGGGACAGTAGGGGAGCGCTG 483  
 Qy 142 LeuSerProArgProLysGlySerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 Db 484 CTCCTCCCGAGGCTGCTCTACTTGAAGGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543  
 Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGTGCGGTATGCACCGGGGGGTTCGGAAGCG 603  
 Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 Db 604 GTGGACTTTGTCCCGTAGAGTCCATGGAACACTATGCGGTCTCCG 651

RESULT 12  
 AAX80354  
 ID AAX80354 standard; cDNA; 1998 BP.  
 XX AC AAX80354;  
 XX DT 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX OS Hepatitis C virus.  
 XX OS Synthetic.  
 XX PN W09928482-A2.  
 XX PD 10-JUN-1999.

```

PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 163-166; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:
Pred. No.: 3.34e-73 Length: 1998
Score: 881.50 Matches: 168
Percent Similarity: 92.86% Conservative: 14
Best Local Similarity: 85.71% Mismatches: 11
Query Match: 86.42% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80354 (1-1998)
OY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTCTGTTATGTTGTTAGATATTTATCTGGTAGTGTGTATCATCGGCTAC 123
OY 22 AlaGlnGlnThrArgGlyGluGlnGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGCCCTACTTGGTTCATCAAGACTAGCCTTACAGCGCGGCAAG 183
OY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACGAGTCGAGGAGAGGTTCCAGTGGTTCCACGCCACACAACTCTCTCGGCGACC 243
OY 62 SerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGCCTCAACGGCGTGTGTTGGACCGTTTACCATGTTGGCTGGCTCAAGACCTTAGCCGC 303
OY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCGCAATCACCCAGATGTACACTAATGTGGACCGACCTCGTGGCTGGCAG 363
OY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuLysLeu 121
DB 364 CGCGCCCCCGGGCGCGCTTCTTGACACCATGACACCTGTGGCAGCTCAGACCTTTACTTG 423
OY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 CTCACGACATGTGACGTGATTCCTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 483
OY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCCAGCCCTGCTCTCTACTTCAAGGGGCTCTTCGGGTGGTCCACTGCTCTCCCT 543
OY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

```

```

Db 544 TCGGGCAGCGTGTGGGCATCTTCGGGGTCCGCTATGCACCGGGGGTTCGAGGCG 603
OY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTGTGCGCGTAGAGTCCATGGAAGTACTATGCGGCTCTCCG 651

RESULT 13
AAX80345
ID AAX80345 standard; cDNA; 651 BP.
XX
AC AAX80345;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
XX
PN W09928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98WO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
DR WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 147-148; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;

Alignment Scores:
Pred. No.: 1.58e-73 Length: 651
Score: 878.50 Matches: 167
Percent Similarity: 93.33% Conservative: 15
Best Local Similarity: 85.64% Mismatches: 10
Query Match: 86.13% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80345 (1-651)
OY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTCTCTGTTATGTTGTTAGATATTTATCTGGTAGTGTGTATCATCGGCTAC 123
OY 22 AlaGlnGlnThrArgGlyGluGlnGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41

```

```
Db 124 TCCCAACAGCGGGCGCTACTGTGTCAGAGACACTAGCTTACAGCGCGGACACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 ACCAGGTGAGGAGAGGTTGAGGTGTTCCACCGCAACACATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGCGGTGTGTGGACCGTTTACCATGTGTGCTGCGTCAAGAGACCTTAGCGGCG 303
Qy 82 ProLysGlyProValThrGlnMetTyThrAsnValAlaPlysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATACCCAGATGTACATAATGTGGACACGAGACCTCTCGGCTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyLeu 121
Db 364 GCGCCCCCGGGCGGTCTCTGACACCATGCTGCGGAGCTGACACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTCTATCCGGTCCGCGCGCGGCGGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCAGGCGCTGTCTCTACTTTGAAGGGCTCTTCGGGTGGTCCACTGCTGCGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGCAGCGTGTGGCATCTTCGGGCTGCGGTATGCACCGGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 604 GTGACATTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCT 648
```

## RESULT 14

```
AA80358
ID AAX80358 standard; cDNA; 1998 BP.
XX
AC AAX80358;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:108.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
XX
XX 10-JUN-1999.
XX
XX 24-NOV-1998; 98WO-US24528.
XX
XX 28-JUL-1998; 98US-0094331.
XX
XX 28-NOV-1997; 97US-0067315.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
XX WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 176-179; 21pp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
```

```
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
```

Sequence 1998 BP: 410 A; 596 C; 568 G; 424 T; 0 other;

## Alignment Scores:

Pred. No.:	6,39e-73	Length:	1998
Score:	878.50	Matches:	167
Percent Similarity:	92.86%	Conservative:	15
Best Local Similarity:	85.20%	Mismatches:	11
Query Match:	86.13%	Indels:	3
DB:	20	Gaps:	1

US-09-965-594-20 (1-197) x AAX80358 (1-1998)

```
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTy 21
Db 64 GGTTCGTGTGTATTGTTGGTAGAATTATTTATCTGGTAGTGTATATCAGCGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGGCTACTTGTGTCATCAGACTAGCCTTACAGCGCGGACAG 183
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACAACTCTCTCGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGCGGTGTGTGTGGACCGTTTACCATGTGTGCTCAAGACCTTAGCGGCG 303
Qy 82 ProLysGlyProValThrGlnMetTyThrAsnValAlaPlysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATACCCAGATGTACATAATGTGGACACGAGCTGCTGGTGGCAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyLeu 121
Db 364 GCGCCCCCGGGCGGTCTCTGACACCATGCTGCGACCTGCGCAGCTCAGACCTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCACGAGACATGCTGACGTCTATCCGGTCCGCGCGGCGGCGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCCTCCCGCAGGCGCTGTCTCTACTTTGAAGGGCTCTGCTGCTGCTCCACTGCTGCGCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGCAGCGTGTGGCATCTTCGGGCTGCGGTATGCACCGCGGGGTTCGGAAGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGACATTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651
```

## RESULT 15

```
AA80353
ID AAX80353 standard; cDNA; 1998 BP.
XX
AC AAX80353;
XX
DT 07-SEP-1999 (first entry)
XX
XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.
XX
```

KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX Hepatitis C virus.  
 OS Synthetic.

XX WO928482-A2.  
 XX 10-JUN-1999.

XX PF 24-NOV-1998; 98WO-US24528.  
 XX PR 28-JUL-1998; 98US-0094331.  
 XX PR 28-NOV-1997; 97US-0067315.  
 XX PA (SCHE ) SCHERING CORP.

XX PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX WI: 1999-385385/32.  
 XX New hepatitis C virus covalent complexes

PS Disclosure: Page 160-162; 21pp; English.  
 XX

CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

XX Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

Alignment Scores:  
 Pred. No.: 7,93e-73 Length: 1998  
 Score: 877.50 Matches: 167  
 Percent Similarity: 92.86% Conservative: 15  
 Best Local Similarity: 85.20% Mismatches: 11  
 Query Match: 86.03% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-20 (1-197) x AAX80353 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTCTGTGTTATTGTTGGTAGAATATTTATCTGGTAGTGGTAGTATCATCGGCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACCGCGGGCTACTTGGTGGCAAGATCATTAGCTTACAGCGCGGACAG 183  
 QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTTCAGGGAGAGGTTTCAGGTGGTTTCCACCGCAACAACATCTCTCTGGCGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGGTCAACGGCGTGTGTGGACCGTTTACCATTGGTGGCTCAAGACCTTAGCCGGC 303  
 QY 82 ProLysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTrpGln 101  
 DB 304 CCNAGGGGCCAATCACCAGATGTACTAATGTGGACAGGACCTCGTCGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerAspLeuTyrLeu 121  
 DB 121

DB 364 GCGCCCCCGGGCGGCTTCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCACGAGACATGCTGACGTCAATCCGGTGCCTCCGCGGGGCGACAGTAGGGGAGCGTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CTCTCCCCCAGGCTGTCTCTCTACTTGAAGGCTCTTCGGGTGGTCCACTGCTCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 544 TCGGGGCACGCTGTGGGCATCTTCCGGGTGCGGTATGCACCGGGGGGTTCGCAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCCCCGTAGAGTCCATGGAACTACTATGCGGTCTCCG 651

Search completed: August 30, 2003, 19:48:15  
 Job time : 188.939 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-20  
Perfect score: 1020  
Sequence: 1 MKKGSVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-Q/cgn2.1/USPTO\_spool/US09965594/runat\_29082003\_151919\_28322/app\_query.fasta\_1.2872  
-DB-EST -OFMT-fastap -SUFFIX-rst -MINMATCH-0.1 -LOOPEL-0 -LOOPEXT-0  
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40 cdi -LIST-45  
-DOCLIGN-200 -THR\_SCORE-pct -THR\_MAX-100 -THR\_MIN-0 -ALIGN-15 -MODE-LOCAL  
-OUTFMT-pco -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000  
-USER-US09965594@cgn\_1.12830.#runat\_29082003\_151919\_28322 -NCPU-3  
-NO\_MAP -LARGEQUERY -NEG\_SCORE-0 -WAIT -DSPBLOCK-100 -LONGLOG  
-DEV\_TIMEOUT-120 -WARN\_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FCGAPOP-6  
-FCGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database : EST:.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_nam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106.5	10.4	1031	14	CB950999
2	106	10.4	1146	12	BM915803
3	102.5	10.0	1403	13	BQ926101
4	101	9.9	1199	13	BQ892487
5	97.5	9.6	629	10	BG089727
6	97	9.5	644	29	BX238988
7	95.5	9.4	701	10	BF863244
8	95.5	9.4	984	10	BF304699
9	94	9.2	560	28	A0538021
10	94	9.2	935	10	B868757
11	94	9.2	1062	29	CHS060HN
12	93.5	9.2	772	29	CC406704
13	93.5	9.2	789	29	CC406705
14	93	9.1	528	12	BM402566
15	93	9.1	701	29	B2342381
16	93	9.1	1213	13	B0541777
17	92.5	9.1	938	13	B0894657
18	91.5	9.0	502	9	AA036834
19	91.5	9.0	528	28	A0620249
20	91.5	9.0	812	13	B0299264
21	91.5	9.0	817	13	B0240438
22	91.5	9.0	878	13	B0365755
23	91	8.9	574	29	CG380642
24	91	8.9	579	29	CG380645
25	91	8.9	580	14	CAT28398
26	91	8.9	753	13	B0402910
27	91	8.9	866	13	B0219343
28	91	8.9	905	13	B0542842
29	91	8.9	906	13	BX434207
30	91	8.9	936	29	CG373208
31	90.5	8.9	622	9	AV835401
32	90.5	8.9	814	11	CHS09179
33	90.5	8.9	824	13	B0396924
34	90.5	8.9	958	10	BG420860
35	90.5	8.9	1035	10	BE888775
36	90.5	8.9	1141	11	AK080545
37	90.5	8.9	1440	12	BM467279
38	90.5	8.9	1733	12	BM553374
39	90.5	8.9	3215	11	AK051518
40	90	8.8	500	12	BM708007
41	90	8.8	569	12	BM825317
42	90	8.8	617	10	BE05938
43	90	8.8	631	10	AK961059
44	90	8.8	658	12	BM830847
45	90	8.8	757	12	B1258851

ALIGNMENTS

RESULT 1  
CB950999  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

CB950999 1031 bp mRNA linear EST 29-APR-2003  
AGENCOURT\_13445496 NIH\_MGC\_177 Mus musculus CDNA clone  
IMAGE:30316162 5', mRNA sequence.  
CB950999  
CB950999.1 GI:30205777  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1031)

DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDCM107 row: b column: 11  
 High quality sequence stop: 333.

Email: cgrabb@remail.nih.gov  
Tissue Procurement: DCTD/Drp  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution by: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

```

/clone_id="RA602.5462030
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 41"

```

Qy	22	AlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSer-HisThrGlyArgAsp	41
Db	1098	GGCGAGACGGGTGTCGTGGAGGAGGTGGCTCCGGCTCTCGCTACCGTCGGTGGAG	1039
Qy	41	SasGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla	61
Db	1038	CTGGAGGACAGAGGT-----CTACGGCGTGGGTAGGGA	1003

```

QY 61 rSerIleAsnGlyValLeuTrpThrValTyrHis-----GlyAlaGlyThr-- 76
Db 1002 CCGCGCTGGTGGATGTTGTGG-----TATCACTTCCCGCGCGGGGAGGAGTACGTG 949

QY 77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90
Db 948 ACGGAGGGCGCCCGCTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903

QY 90 rThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGln-----GlySerAr 107
Db 902 -----CAGATGTGCGGTGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 859

QY 107 gSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAs 127
Db 858 ACTTCTGTCTGCTGTTTCTGTGGG-----GlySerProValThrGlnMetTy 834

QY 127 pValIleProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProI 147
Db 833 -----CGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787

QY 147 eSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValG 167
Db 786 CCGGTATCTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742

QY 167 yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IlePr 186
Db 741 GCGCTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 682

QY 186 oVal 187
Db 681 TTTG 678

RESULT 3
BQ926101/c
LOCUS BQ926101.1 1403 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION BQ926101
VERSION BQ926101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1403)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AMI3798 row: j column: 07
High quality sequence stop: 101.
Location/Qualifiers
1. 1403
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/notes="Organ: Oocytes; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a

```

```

NIH_MGC Library..
BASE COUNT 297 a 521 c 237 g 345 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 8.34 Length: 1403
Score: 102.50 Matches: 58
Percent Similarity: 36.00% Conservative: 14
Best Local Similarity: 29.00% Mismatches: 66
Query Match: 10.05% Indels: 62
DB: 13 Gaps: 11

US-09-965-594-20 (1-197) x BQ926101 (1-1403)
QY 11 GlyArgIleAsnLeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGly 30
Db 1378 GGTGTGTTTCANCGGTTCAGGACAGGTGCGC---GCACACTCGACCGCTCGGCAGAGACT 1322

QY 31 CysGlnLysThr-----SerHisThrGlyArgAspLysAsnGln-----Val 44
Db 1321 TGTCTGGGGCGCGCTTGGCGCATACCGCGGTGCGGTGAGGTGAGGTGAGGTGAGGTGAG 1262

QY 45 GluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsn 64
Db 1261 GAGGGGAAA-----CAG 1250

QY 65 GlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys--G 84
Db 1249 GGGGTA---TGGTTATCAGCGGCTGGGCGAGGTACT-----TCCCTAAAGCG 1205

QY 84 lyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProG 104
Db 1204 CGCGCTGGCGAGTATATATACCGCGGAGTGGCGAAGCGCACGGGCGGTGGACGTTGACC 1145

QY 104 lnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrA 124
Db 1144 AA---GAGAGCGACCTGACCGCCCTCCCTGTGGGCTGTCGATATACAAATATGTCAG 1088

QY 124 rgHisAlaAspValIleProValArgArgGlyAsp-----136
Db 1087 GGCACGGTGTGTTGTTACTACGCGCGAGTGGCGACCGCGCTCCACACGGCGCTCTTACAG 1028

QY 137 -----SerArgGlySerLeuSerProArgProIle-SerTyrLeuLysGlySerSer 154
Db 1027 CCGCTCCCGCGGCAAC-----AGGTAATATCATATGCGGCGGGGATTC 980

QY 155 Gly-----GlyProLeuLeuLys 160
Db 979 GCATTCGCGGGGAGAGCGCGGTCTCGGGGCGCGCGCTCGCGCGCGCTGAGCGCG 920

QY 161 ProAlaGlyHisAlaValGlyTyrPheArgAlaAlaValSerThrArgGlyVal 178
Db 919 AGGAGAGGC-----GGCGTGTTCGGCGGTGAGGACGAGGCGCGCGGTG 875

RESULT 4
BQ92487
LOCUS BQ92487
DEFINITION AGENCOURT_8417538 Lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6192708 5', mRNA sequence.
ACCESSION BQ92487
VERSION BQ92487.1 GI:22284501
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1199)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski

```

CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MCC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L14M13595 row: c column: 13  
 High quality sequence start: 57  
 High quality sequence stop: 394.

## FEATURES

source  
 1. .1199  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6192708"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski\_sympathetic\_trunk"  
 Note: Site 2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACGACGCGGCGG-3' and  
 5'-GACGAGTCTGATGCGGCGGCGGCTT(15)-3'. Size selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine); available through Life  
 Technologies.  
 BASE COUNT 255 a 362 c 343 g 211 t 28 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.59 Length: 1199  
 Score: 101.00 Matches: 50  
 Percent Similarity: 34.74% Conservative: 24  
 Best Local Similarity: 23.47% Mismatches: 80  
 Query Match: 9.90% Indels: 59  
 DB: 13 Gaps: 9

US-09-965-594-20 (1-197) x BQ892487 (1-1199)

QY 16 SerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSer 35  
 DB 337 CAGAGAGAGAACCTTACCCCAACAG-----AAGGCA 369  
 QY 36 HistHrGlyArgAspLysAsnGlnValGluGlyValGlnLysValSerThr----- 53  
 DB 370 CATGGGGAAATCGCCGCTTCAGAGACGAGGTTCATGTTTCTGAAACATAACCG 429  
 QY 54 AlaThrGlnThrPheLeu-----AlaThrSerIleAsnGlyValLeuThrThr 69  
 DB 430 CAGCCACTGTCTTCATGTAATGATACCCCTTCCACCACACACAGGGCAGCAGCGATGGAT 489  
 QY 70 ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet 89  
 DB 490 CCAATTTTAAAGGGTGCTCTGTTAATCATGCGGCCCGCCGCGTACATCTCCA 549  
 QY 90 TyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeu 109  
 DB 550 TTTACCACATGTGACGTGACTTT-----TGT 576  
 QY 110 ThrProCysThr-----CysGlySerSerAspLeuTyr 120  
 DB 577 GTGCTCTGCACAGACACCCCGATGACCGATGTGGCTTATGTGAACGGCGGCGGCTC 636  
 QY 121 -LeuValThr-----ArgHisAlaAspValIleProValArg----- 132  
 DB 637 ATTGGCCACTCCCTCTTATAAACACAGCCAGCTGCTCCATGGCGGGCGTGGGTGT 696  
 QY 133 -----ArgArgGlyAspSerArgGlySerLeuLeu----- 142

DB 697 TTGCAGCGCAAGCGGGTGGGGCTAGGACTCGGGGGCGGATTCTCTGAAACC 756  
 QY 143 -SerProArgProIleSerTyrLeuLys-----GlySerSerGlyGlyPr 157  
 DB 757 CCACCTCGGGCCACCGCATGCGCTTAAGCCTCCCTTTTAAAGCCACCGCGCGGCCCC 816  
 QY 157 OleuLeuLysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgG 177  
 DB 817 CTTACATCTCTACCTCGGGCGGGGGGAGAGACGTGGCGGCATACGGCTCAGGG 876  
 QY 177 yValAlaLysAlaValAspPheIleProValGluSer 189  
 DB 877 CGTTTAAAGCCCGCGCTTCGCGCGGGCGGAAGCA 913

## RESULT 5

BQ89727/c  
 LOCUS  
 DEFINITION  
 BQ89727 629 bp mRNA linear EST 26-JAN-2001  
 similar to SW:GRAD\_MOUSE P11033 GRANZYME D PRECURSOR ;, mRNA  
 sequence.  
 ACCESSION  
 BQ89727  
 VERSION  
 BQ89727.1 GI:12572290  
 KEYWORDS  
 EST.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 629)  
 AUTHORS  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

MGI:1477610

Seq primer: -40UP from Gibco  
 High quality sequence stop: 422.

## FEATURES

source  
 1. 629  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3977578"  
 /tissue\_type="NK cells (flow-sorted)"  
 /lab\_host="DH10B (TI-resistant)"  
 /clone\_lib="NCI-CGAP\_Sp2"  
 /note="Organ: spleen; Vector: pCMV-SPORT6 (Life  
 Technologies); mRNA made from flow-sorted NK cells, CDNA  
 made by oligo-dT priming. Directionally cloned. Average  
 insert size 1.5 kb. Primary library, non-amplified. CDNA  
 Library Preparation: David B. Krizman, Ph.D."  
 BASE COUNT 131 a 156 c 150 g 191 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 9.21 Length: 629  
 Score: 97.50 Matches: 48  
 Percent Similarity: 39.23% Conservative: 23  
 Best Local Similarity: 26.52% Mismatches: 59  
 Query Match: 9.56% Indels: 51  
 DB: 10 Gaps: 12

US-09-965-594-20 (1-197) x BQ89727 (1-629)

QY 36 HistHrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThr 55

```

Db 620 CATCCGGTAAG-----GAAGGAGACACACAGATCATCCCTTGTGCA--- 579
QY 56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrValThrHisGly----- 73
Db 578 AAACATTTCCCATCCAGATATAAATGCT-----ACTATCTCTTCAGGTGAGATC 528
QY 74 -----AlaGlyThrArgThrIleAlaSer 81
Db 527 ATGCTGTTAAAGCTGGAGAGTAAGCCCAAGAGAACTAAAGCTGTGAGACCCCTCAAGTTG 468
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
Db 467 CCGAGATCCAAATGCCCGGCTGAAGCCAGGAATGTG---TGCAGTGTGCTGGCTGG--- 414
QY 102 AlaProGlnGlySerArgSerLeu-----ThrProCysThrCysGlySerSerAspLeu 119
Db 413 -----GGTCAAGTCTCATCAATACACATCAAGCATCTGCCGCTCCGAGAGGTT 363
QY 120 TyrLeuValThrArgHisAlaAspValIleProValArgArgArg----- 134
Db 362 CAACCTGGTTCATCCAGGAGGAGGAGGAGGATGCAAAAAACGTTTCCGATACACTAGACAC 303
QY 135 -----GlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr 149
Db 302 ACAGAGATTGTGCTGGAGACTTGAAG---AAAAATNAAGACTCT----- 261
QY 150 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169
Db 260 TTCAGGGTGTACTCGGGGGGACCCCTTGTGTGTGAC---AACCAAGCATATGACTTTTC 204
QY 170 ArgAlaAla-----ValSerThrArgGlyValAlaLysAlaValAspPheIle 185
Db 203 GCCTATGCAAAAAACGGAACAATCTCTTCAGGAATCTTCACTAAGTTGTGCACCTTCCTG 144
QY 186 Pro 186
Db 143 CCG 141

RESULT 6
LOCUS BX238988 644 bp DNA linear GSS 29-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-283L13, genomic survey sequence.
ACCESSION BX238988
VERSION BX238988.1 GI:28161322
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 644)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 283L13. 283L13
is part of the Daniokey BAC Library created by R. Plasterk and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_danio/.
FEATURES
source
1..644
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-283L13"
/tissue_type="Testis"
/notes="vector pIndigoBAC-536"
BASE COUNT 129 a 212 c 176 g 127 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 10.7 Length: 644
Score: 97.00 Matches: 48
Percent Similarity: 41.95% Conservative: 25
Best Local Similarity: 27.59% Mismatches: 84
Query Match: 9.51% Indels: 17
DB: 29 Gaps: 7

US-09-965-594-20 (1-197) x BX238988 (1-644)
QY 18 AspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThr 37
Db 556 GACCATCGCTATTACCATCATCTGGAGGGGAGACAGCTCGAATCCGGACGGCCATCTC 497
QY 38 GlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThr 57
Db 496 CTTGACGGGTGACGCCAGCCGTCCTCCCTGGCCGGGTGCAGAGTCCCATAGGGCC 437
QY 58 PheLeuAlaThrSerIleAsnGlyValLeu-TripThrValTyrHisGlyAlaGlyThrAr 77
Db 436 GAGACTGCATGGACCCACCGCGTCCGACAGCATCTCTGGCAGAGTGGGAGCTC 377
QY 77 gThrIleAlaSerProLysGlyProValThrGlnMetTyr-----ThrAs 92
Db 376 TCGAGGCGCTGTCCGCGCTCACAGTGCACGACGAGCTGGTCATGTGGTGTGTCACGAC 317
QY 92 nValAspLysAspLeuValGlyTyrGlnAla----ProGlnGlySerArg-----Se 108
Db 316 TCITGGAAAGGAC-----TGGAGATCCACACCGTTGGGGACAAAGCGGAGAGTCT 266
QY 108 rLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspVa 128
Db 265 GTCACCAACCATGCCCTTTAAAAAAGGAGAAATATTATTGATTACCATGGGAGGAGTTG 206
QY 128 lIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgPro----- 146
Db 205 TCGCAGCAGAGGGCTTAGGGAGACGGGAGGCCCTCCCGCTGTCTCCTCTCATATG 146
QY 147 -lIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuLeu---CysProAlaGlyHisAl 165
Db 145 TATTCTCT---TTAAAGGCTTGTGGGAGGACCCCTTTCCTGGATGCCAGTCTCGGCC 89
QY 165 aValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
Db 88 TGCCCTCTGCACCCGGGCATGGAACCTTCGGAAGGCTTA 49

RESULT 7
LOCUS BF863244 701 bp mRNA linear EST 19-JAN-2001
DEFINITION 963042C02.x1 C. reinhardtii CC-1690, Stress condition 1, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF863244
VERSION BF863244.1 GI:12253388
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE 1 (bases 1 to 701)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..701

```

/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 2lgr"  
/db\_xref="taxon:3055"

/clone\_lib="c. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with EXAssist (Stratagene) phage. The library was normalized with method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 173 a 213 c 175 g 140 t

## ORIGIN

Alignment Scores:  
Pred. No.: 16.8 Length: 701  
Score: 95.50 Matches: 38  
Percent Similarity: 38.96% Conservatives: 22  
Best Local Similarity: 24.68% Mismatches: 63  
Query Match: 9.36% Indels: 31  
DB: 10 Gaps: 7

US-09-965-594-20 (1-197) x BF863244 (1-701)

QY 71 TyrHisGlyAlaGlyThrArgThrIleAlaSerProLys-----GlyProVal 86  
Db 171 CACCACCATACCCCTGCTCTCAGCTGCTCACACCAAAATATGCCATACGGGCGCACTA 230  
QY 87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySer 106  
Db 231 ACAAGTTTACATACACGG-----AAGCACCGCGCTTGGCCACCCCTTGGAGCGG 284  
QY 107 ArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAla 126  
Db 285 AGAAGCCGCGCTGCTCTGGGTCATCCGCATGCTATGCAATCTCCCGCTATCAG 344  
QY 127 AspValIle-----ProValArgArgArgGlyAspSerArg----- 138  
Db 345 GAGATCATTTGTCATGTGGCTTTAGTCAACCCCAAGAGAGCGCTGGGAGTGGCATTTATAA 404  
QY 139 -----GlySerLeuSerProArgProLysSer---Tyr 149  
Db 405 GAAGGGNCGGAATTCGGTTTCGGAAGAGTGAAGCGCCCAAGGCTGACCAAGTGCTA 464  
QY 150 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe 169  
Db 465 CTCGAAGCGAGCAATGGGAGCGTTTCGCGGTGTCGCGGTCTGCTCCTCTAATGTGACG 524  
QY 170 ArgAlaAlaVal-----SerThrArgGlyValAlaLysAla--- 181  
Db 525 AAAGAACCATTGAGTAGGAAGTGGCGGTTTACCCCGCAAGGTGAAGTCACTCTAT 584  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArg 195  
Db 585 GTGAGCCGCATACATCTGGAGACACACGCGTGACACTCTACGA 626

RESULT 8  
BF304699/c 984 bp mRNA linear EST 21-NOV-2000  
LOCUS 601888252F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:412276 5',  
DEFINITION mRNA sequence.  
ACCESSION BF304699  
VERSION BF304699.1 GI:11251586  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 984)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)

Plate: LLCM1005 row: g column: 13

High quality sequence stop: 646.

FEATURES Location/Qualifiers

1..984

1. .984

Location/Qualifiers

1. .984

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers



Qy 18 AspThrAlaTyraAlaGlnGlnThrArgGlyGluGlnGlyCysGlnLysThrSerHisThr 37  
 Db 611 GCCAAACATACACGGACCAACAGATACGGTACAGGGTGCCAAAG----- 658  
 Qy 38 GlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThr 57  
 Db 659 GGAGCGGATAAAGGCTAGGCAACGGG-----CCCGGAGTAACCC 700  
 Qy 58 PheLeuAlaThrSerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArg 77  
 Db 701 GCGTCGGCGCGGGGACACGGGGAAACCCCGGGGAAACCCCTCGGACAGGGAACGCTG 760  
 Qy 78 ThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeu 97  
 Db 761 ACAAAACCGCGGAGGAGGACCTCAACGCGGCCCCACACCAACCTGACCGCAACAT- 819  
 Qy 98 ValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSer 117  
 Db 820 -----TACACGCCCCACACGGG-----ACACCACTA----- 846  
 Qy 118 AspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSer 137  
 Db 847 CACAAATTCACACGGCGGCTCACCCCGGGTCCACCCTACGGCGGAGAAATCATCAGCGC 906  
 Qy 138 ArgGlySer 140  
 Db 907 CGAGCGTCA 915

## RESULT 11

CNS06QHN 1062 bp DNA linear GSS 05-JUL-2001  
 LOCUS T3 end of clone AWOAA006B03 of library AWOAA from strain CLIB 89 of  
 Yarrowia lipolytica, genomic survey sequence.  
 ACCESSION AL410673  
 VERSION AL410673.1 GI:12179275  
 KEYWORDS GSS.  
 SOURCE Yarrowia lipolytica  
 ORGANISM Yarrowia lipolytica  
 Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Dipodascaceae; Yarrowia.  
 REFERENCE 1 (bases 1 to 1062)  
 AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,  
 de-Montigny,B., DuJon,B., Durrens,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Wincker,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 1062)  
 AUTHORS Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,  
 Artiguenave,F., Wincker,P. and Gaillardin,C.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia  
 lipolytica  
 JOURNAL FEBS Lett. 487 (1), 95-100 (2000)  
 MEDLINE 20584727  
 PUBMED 11152892  
 REFERENCE 3 (bases 1 to 1062)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,  
 2 rue Gaston Cremieux, Cp 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
 seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

## FEATURES

source location/Qualifiers  
 1..1062  
 /organism="Yarrowia lipolytica"  
 /mol\_type="genomic DNA"  
 /strain="CLIB 89"  
 /db\_xref="taxon:4952"  
 /clone="AWOAA006B03"  
 /clone\_lib="AWOAA"  
 /note="end : T3"  
 /note=">917  
 PTR2 : peptide transporter ]"  
 /evidence=not\_experimental  
 misc\_feature  
 274 a 264 c 242 g 280 t 2 others  
 BASE COUNT  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 40.8 Length: 1062  
 Score: 94.00 Matches: 50  
 Percent Similarity: 36.10% Conservative: 24  
 Best Local Similarity: 24.39% Mismatches: 57  
 Query Match: 9.22% Indels: 74  
 DB: 29 Gaps: 11  
 US-09-965-594-20 (1-197) x CNS06QHN (1-1062)

Qy 32 GlnLysThrSerHisThr-----GlyArgAspLysAsnGlnValGlu-----GlyGluVal 48  
 Db 447 CAAGAAGCTACCACTCTATCGATAAGACGACAAAAGAAATGAATTCACCGAGATT 506  
 Qy 49 GlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTyr 68  
 Db 507 GAACACATCGATCATCTCCCGGGTACTCTCAGACA-----TGG 548  
 Qy 69 ThrValTyr-----HisGly----- 73  
 Db 549 GCCACTACACCGGATGAACATACCCCGCGGCTCAGAATTGCCACAGAGAAATCGA 608  
 Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet----- 89  
 Db 609 GCGACTTAAGACGAGTAGCTCTCCAAATGGAGCCATGACTTACATGCTCTCTTGT 668  
 Qy 90 -----TyrThrAsnValAspLysAspLeuValGly 99  
 Db 669 GAGTTCGCAGACGAGGCTCGTACTATGCTGACCAACGTCATTTCCAACTTTGTCCAG 728  
 Qy 100 TrrGlnAlaProGlnGlySerArg-----SerLeuThrProCys-----Thr 113  
 Db 729 TTCCCTCTCCCTAAGGGCGGAATGGTGGGAGCCACACCTCGCGGTTCACATTGACA 788  
 Qy 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArg 133  
 Db 789 GCGGAGCCCTAGATCAGGGTCTTCAAGTCGCCAGGCTCTGACTCTGGTTTTCAGTTT 848  
 Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153  
 Db 849 TTA-----TCCTACTTGACCTCTTCTAGGAGCCTACCTTCCCGATTC 893  
 Qy 154 Ser-----GlyGlyProLeuLeuCysProAlaGlyHis--- 164  
 Db 894 AAATATGGACGTTTCAAGACCATCTGGCGCGGTACGATCATCTGTGGNAATGGCCATTT 953  
 Qy 165 -----AlaValGlyIle 168  
 Db 954 GTGATTGTGATTCGCGGAATTCGCGGAATCATAGACACAGAAAGCTGCTCTAGGAATC 1013  
 Qy 169 PheArgAlaAlaVal 173  
 Db 1014 TTTATTGCTGGACTA 1028



```

RESULT 12
CC406704/c
LOCUS          772 bp      DNA      linear      GSS 19-MAY-2003
DEFINITION    PUHKL12TB 2M_0.6_1.0_KB zea mays genomic clone ZMMBta469B24,
               genomic survey sequence.
ACCESSION     CC406704
VERSION       CC406704.1 GI:30886794
KEYWORDS      GSS.
SOURCE        Zea mays
ORGANISM      Zea mays
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
               clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 772)
AUTHORS       Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
               A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE         Maize Genomics Consortium
JOURNAL       Unpublished
COMMENT       Other_GSSs: PUHKL12TB
               Contact: Cathy Whitelaw
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301-838-5843
               Fax: 301-838-0208
               Email: whitelaw@tigr.org
               Seq primer: TR
               Class: sheared ends.
               Location/Qualifiers
                 1..772
                   /organism="Zea mays"
                   /mol_type="genomic DNA"
                   /strain="B73"
                   /db_xref="taxon:4577"
                   /clone_lib="ZMMBta469B24"
                   /clone_lib="2M_0.6_1.0_KB"
                   /note="vector: PCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
                   Cot selected genomic DNA library"
BASE COUNT    150 a 198 c 216 g 208 t
ORIGIN
Alignment Scores:
Pred. No.:      30.1      Length:      772
Score:          93.50     Matches:      49
Percent Similarity: 39.46% Conservative: 24
Best Local Similarity: 26.49% Mismatches: 67
Query Match:     9.17%   Indels:      46
DB:              29      Gaps:       10

US-09-965-594-20 (1-197) x CC406704 (1-772)
Qy 29 GlnGlyCysGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
Db 663 CAGTCGTTCAAGCGCCTCACCCCGCGAGATGGCGGAACGACGTAACAAGGGCTAT 604
Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66
Db 603 GTTACAAATGTGAGC-----AGCCCTATGTAGAGGCCACCGCTGTCCAGGCTATTCT 550
Qy 67 LeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
Db 549 ATCTGGAGG-----TCACAGACTTTGCTGACGATGAGATCAAGGGTG 508
Qy 87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp-----GlnAlaPro 103
Db 507 ACACAGATGATGACAGG-----AAGAGCAGGACCCAGTGTTCCTACATGCAATAA 454
Qy 104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
Db 453 CTGGGATTAGAGGGAGGACCATCAGCTGCGGTAGTCTCTCAACGTCACAGGAGCTGC 394
Qy 120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
Db 393 TGGCCCTACTTGACCGGTTCAACACATAACTTCATCACTCAACAGGGCGCA-CNACAG 335

```

---

```

Qy 138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyPro 157
Db 334 CGTGGGTTACTTTGGAAACCCACACAGGTGCGCATGTCGAAGTGGCAATATGGACACCA 275
Qy 158 LeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
Db 274 GTTTTCTGCCAG-----GGAGTAACCTGTCGCGCA----- 245
Qy 178 ValAlaLysAlaValAspPhe-----IlePro 186
Db 244 -----GCCATTGACATCAACAGGAGAGTTCACCATGAGGCATATGCAATTCCC 194
Qy 187 ValGluSerLeuGlu 191
Db 193 TTGATACATTGTAG 179

RESULT 13
CC406705
LOCUS          789 bp      DNA      linear      GSS 19-MAY-2003
DEFINITION    PUHKL12TD 2M_0.6_1.0_KB zea mays genomic clone ZMMBta469B24,
               genomic survey sequence.
ACCESSION     CC406705
VERSION       CC406705.1 GI:30886795
KEYWORDS      GSS.
SOURCE        Zea mays
ORGANISM      Zea mays
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
               clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 789)
AUTHORS       Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
               A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE         Maize Genomics Consortium
JOURNAL       Unpublished
COMMENT       Other_GSSs: PUHKL12TB
               Contact: Cathy Whitelaw
               TIGR
               9712 Medical Center Drive, Rockville, MD 20850, USA
               Tel: 301-838-5843
               Fax: 301-838-0208
               Email: whitelaw@tigr.org
               Seq primer: TF
               Class: sheared ends.
               Location/Qualifiers
                 1..789
                   /organism="Zea mays"
                   /mol_type="genomic DNA"
                   /strain="B73"
                   /db_xref="taxon:4577"
                   /clone_lib="ZMMBta469B24"
                   /clone_lib="2M_0.6_1.0_KB"
                   /note="vector: PCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
                   Cot selected genomic DNA library"
BASE COUNT    210 a 220 c 203 g 156 t
ORIGIN
Alignment Scores:
Pred. No.:      31      Length:      789
Score:          93.50     Matches:      49
Percent Similarity: 39.46% Conservative: 24
Best Local Similarity: 26.49% Mismatches: 67
Query Match:     9.17%   Indels:      46
DB:              29      Gaps:       10

US-09-965-594-20 (1-197) x CC406705 (1-789)
Qy 29 GlnGlyCysGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
Db 110 CAGTCTGTTCAAGCGCCTCACCCCGCGAGATGGCGGAACGACGTAACAAGGGCTAT 169
Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66

```

```

Db      170 GTTACAATTGTGACG-----AGCCCTATGTACGAGGCCACCGCTGCCAGGCTATTCT 223
QY      67 LeuTTPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
      ::::|||||
Db      224 ATCTGGAGG-----TCACAGACTTGTGACGATGACATCAAGGTTG 265
QY      87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr-----GlnAlaPro 103
      |||||
Db      266 ACACAGATGATGACCAGG-----AAGAGCAGGAGCCAGTGGTTCCCTACATGCAATAA 319
QY      104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
      |||||
Db      320 CTGGGATTAGAGGAGGAGGACCATGCGAGTGGCGGTAGTCTCAACGGTCAGGAGTGTC 379
QY      120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
      |||||
Db      380 TGGCCCTACTTGACACGGTTCACACATCACTTCATCACTGCAAGCGGSCA-CAACAG 438
QY      138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyPro 157
      |||||
Db      439 CTGGGGTTACTTTGGAAACCCACACAGGTCGCCATGTCAAGGTGGCAATGGAGACCCA 498
QY      158 LeuLeuSerProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
      ::::
Db      499 GTTTCTGCCAG-----GGAGTACTGCTCCGCA----- 528
QY      178 ValAlaLysAlaValAspPhe-----IlePro 186
      |||||
Db      529 -----GCCATTGACATCAACAGGAGGAAGTTCAACATTGAGGCATATGCAATTC 579
QY      187 ValGluSerLeuGlu 191
      ::::
Db      580 TTGGATACATTGAG 594

RESULT 14
BM402566
LOCUS      528 bp mRNA linear EST 01-JUL-2002
DEFINITION SLA005f12_34513 An expressed sequence tag (EST) collection from the
            resurrection plant Selaginella lepidophylla
ACCESSION  BM402566
VERSION     1
KEYWORDS   SLA005f12 5, mRNA sequence.
SOURCE     Selaginella
ORGANISM   Selaginella lepidophylla
            Selaginella lepidophylla
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;
            Selaginella.
REFERENCE  1 (bases 1 to 528)
AUTHORS   Iturriaga,G. and Cushman,J.C.
TITLE     An expressed sequence tag (EST) collection from the resurrection
JOURNAL   Plant Selaginella lepidophylla
COMMENT   Unpublished
            Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
PCR PRIMERS
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 005 row: F column: 12
Seq primer: T3 20mer
High quality sequence stop: 528.
Location/Qualifiers
1. 528
/mol_type="mRNA"
/db_xref="taxon:59777"
/clone="SLA005f12"
/tissue_type="microphyll fronds undergoing desiccation for
source

```

```

2.5 h"
/dev_stage="adult"
/clone_lib="An expressed sequence tag (EST) collection
from the resurrection plant Selaginella lepidophylla"
/notes=Vector: Lambda Uni-zap XR, Bluescript SK; Site_1:
EcoRI; Site_2: XhoI; Library construction was performed
according to manufacture's (Stratagene, Inc.) recommended
protocol for the Lambda UniZapXR vector and cDNA synthesis
kit."
BASE COUNT 129 a 125 c 137 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 20.6 Length: 528
Score: 93.00 Matches: 37
Percent Similarity: 42.98% Conservative: 15
Best Local Similarity: 30.58% Mismatches: 43
Query Match: 9.12% Indels: 26
DB: 12 Gaps: 4
US-09-965-594-20 (1-197) x BM402566 (1-528)
QY 94 AspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
      |||||
Db 53 GACAAGATGTAGCGGTGCTGAAGATCGATGCTCAAGCAACAGATCTCAGGCCAATACCC 112
      |||||
QY 114 CysGlySerSerAspLeuTyrLeuVal----- 122
      |||||
Db 113 CTGGAAGTTCGTCCGATCTGCTTGTGGCCAGAGGTGATGCTATCGTAAATCCTTTT 172
      |||||
QY 123 -----ThrArgHisAlaAspValIleProValArgArgGlyAspSerArg 138
      |||
Db 173 GGATTGGATCATACGCTGACACAGCGGCTCATCGTCTTTCGAAGGAGATTTACT--- 229
      |||
QY 139 GlySerLeuLeuSerProArgProIleSerTyrLeu----- 150
      |||
Db 230 ---TCACCGCGCTAATGGTGTCTCAATCCAGACAGTGATCCACAGAGATGCCGCTATTAA 286
      |||
QY 151 LysGlySerSerGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170
      |||||
Db 287 CTGGAACAGCGGGGTCGCTATTTGGACAGTCTCTGGAAATTTGATAGGCATCAACT 346
      |||
QY 171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189
      |||
Db 347 GCTATATATTCCTCGCTGCGGCTTCATCAGCGGTGGGCTTTTCATCCAGTTGACACG 406
      |||
QY 190 Leu 190
      :::
Db 407 GTT 409
      :::
RESULT 15
BZ342381/c
LOCUS      701 bp DNA linear GSS 06-NOV-2002
DEFINITION ic83b11.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
            bicolor genomic clone ic83b11 5', genomic survey sequence.
ACCESSION  BZ342381
VERSION     1
KEYWORDS   BZ342381.1 GI:24742983
SOURCE     GSS.
ORGANISM   Sorghum bicolor (sorghum)
            Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE  1 (bases 1 to 701)
AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimben,L.,
            Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.
TITLE     Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
JOURNAL    Unpublished
COMMENT    Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 44.6227 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVIVGRINLSGDTA.....YAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	100.0	197	AA15225	Hepatitis C virus
2	1010	99.4	197	AA15224	Hepatitis C virus
3	995	97.9	197	AA15223	Hepatitis C virus
4	980	96.5	197	AA15222	Hepatitis C virus
5	963	94.8	197	AA15221	Hepatitis C virus
6	936	92.1	197	AA15226	Hepatitis C virus
7	929	91.4	195	AA15220	Hepatitis C virus
8	902	88.8	195	AA15212	Hepatitis C virus
9	875.5	86.2	665	AA124943	HCV NS4A-NS3 compl

10	872.5	85.9	665	20	AA124947	HCV NS4A-NS3 compl
11	871.5	85.8	665	20	AA124942	HCV NS4A-NS3 compl
12	868.5	85.5	216	20	AA17880	HCV NS4A-NS3 compl
13	868.5	85.5	665	20	AA124946	HCV NS4A-NS3 compl
14	867.5	85.4	665	20	AA124941	HCV NS4A-NS3 compl
15	865.5	85.2	216	20	AA17884	HCV NS4A-NS3 compl
16	864.5	85.1	216	20	AA17879	HCV NS4A-NS3 compl
17	864.5	85.1	665	20	AA124945	HCV NS4A-NS3 compl
18	863.5	85.0	665	20	AA124940	HCV NS4A-NS3 compl
19	863.5	85.0	671	20	AA124948	HCV NS4A-NS3 compl
20	861.5	84.8	216	20	AA17883	HCV NS4A-NS3 compl
21	860.5	84.7	216	20	AA17878	HCV NS4A-NS3 compl
22	860.5	84.7	665	20	AA124944	HCV NS4A-NS3 compl
23	860.5	84.7	671	20	AA124949	HCV NS4A-NS3 compl
24	860	84.6	215	20	AA17890	HCV NS4A-NS3 compl
25	857.5	84.4	216	20	AA17882	HCV NS4A-NS3 compl
26	857.5	84.4	216	20	AA17886	HCV NS4A-NS3 compl
27	856.5	84.3	216	20	AA17877	HCV NS4A-NS3 compl
28	854	84.1	215	20	AA17887	HCV NS4A-NS3 compl
29	853.5	84.0	216	20	AA17881	HCV NS4A-NS3 compl
30	853.5	84.0	216	20	AA17885	HCV NS4A-NS3 compl
31	849	83.6	213	20	AA17888	HCV NS4A-NS3 compl
32	849	83.6	631	20	AA17882	HCV NS3 protein..
33	848.5	83.5	131	21	AA144728	Hepatitis C virus
34	848.5	83.5	3011	19	AA177397	Hepatitis C virus
35	848.5	83.5	3011	24	ABP71460	Amino acid sequence
36	848.5	83.5	3012	23	AA199289	Hepatitis C virus
37	845.5	83.2	3011	14	AA140120	HCV genomic amino
38	844.5	83.1	687	16	AA179223	PHCV150-encoded se
39	844.5	83.1	1648	16	AA179221	PHCV176-encoded se
40	844.5	83.1	1766	10	AA192041	Sequence encoded i
41	844.5	83.1	1766	10	AA190158	Protein sequence o
42	844.5	83.1	2261	10	AA190164	Peptide encoded by
43	844.5	83.1	2301	10	AA192047	Sequence encoded i
44	844.5	83.1	2436	10	AA192050	Sequence encoded i
45	844.5	83.1	2436	10	AA190288	Peptide encoded by

#### ALIGNMENTS

```

RESULT 1
AAB15225
ID  AAB15225 standard; protein: 197 AA.
AC  AAB15225;
DT  19-DEC-2000 (first entry)
DE  Hepatitis C virus NS4A-NS3 fusion protease #7.
XX  Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW  liver failure; liver cancer; mutant; mutein.
XX  Hepatitis C virus.
OS  Synthetic.
PN  WO2000040707-A1.
XX  13-JUL-2000.
XX  06-JAN-2000; 2000WO-US00345.
XX  08-JAN-1999; 99US-0115271.
XX  (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI  Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;
DR  WPL; 2000-465976/40.
DR  N-PSDB; AAA73334.
XX  Modified hepatitis C virus (HCV) NS3 protease comprising at least 1

```

PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 XX  
 PS Claim 23; Fig 17; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 100.0%; Score 1016; DB 21; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-98;  
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGREGTQKTSHTRGRDNQVGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGREGTQKTSHTRGRDNQVGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPTCTGSSDLY 120  
 DB 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPTCTGSSDLY 120  
 QY 121 LVTRHADVIPVRRRGRSGSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRRGRSGSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 2  
 AAB15224  
 ID AAB15224 standard; protein; 197 AA.  
 AC AAB15224;  
 DT 19-DEC-2000 (first entry)  
 XX Hepatitis C virus NS4A-NS3 fusion protease #6.  
 DE  
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 KW  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 DR N-PSDB; AAA73333.  
 XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT -  
 XX  
 XX  
 PS Claim 23; Fig 16; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-7  
 CC variant.  
 XX  
 SQ Sequence 197 AA;

Query Match 99.4%; Score 1010; DB 21; Length 197;  
 Best Local Similarity 99.5%; Pred. No. 9e-98; Mismatches 1; Indels 0; Gaps 0;  
 Matches 196; Conservative 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAAQTGREGTQKTSHTRGRDNQVGEVQIVSTATQTFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAAQTGREGTQKTSHTRGRDNQVGEVQIVSTATQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPTCTGSSDLY 120  
 DB 61 TSINGVLWTVYHGAGTRTIASPKGPVTQMTYNDKDLVGVQAPQGSRLTPTCTGSSDLY 120  
 QY 121 LVTRHADVIPVRRRGRSGSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRRGRSGSLSPRISYLYKSGSGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197

RESULT 3  
 AAB15223  
 ID AAB15223 standard; protein; 197 AA.  
 AC AAB15223;  
 DT 19-DEC-2000 (first entry)  
 XX Hepatitis C virus NS4A-NS3 fusion protease #5.  
 DE  
 DE Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 KW  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 DR N-PSDB; AAA73332.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT  
XX  
PS Claim 23; Fig 15; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
CC variant.  
XX  
SQ Sequence 197 AA:  
Query Match 97.9%; Score 995; DB 21; Length 197;  
Best Local Similarity 98.0%; Pred. No. 3.4e-96;  
Matches 193; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEVQIVSTATQTFLA 60  
DB 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEVQIVSTATQTFLA 60  
QY 61 TSINGVLWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120  
DB 61 TSINGVLWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGPGLPCPAGHAGVIFRAAVSTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGPGLPCPAGHAGVIFRAAVSTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 4  
AAB15222  
ID AAB15222 standard; protein; 197 AA.  
XX  
AC AAB15222;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #4.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX WPI; 2000-465976/40.

DR N-PSDB; AAA73331.  
XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT  
XX  
PS Claim 23; Fig 14; 66pp; English.  
XX  
CC The present sequence is a mutated version of a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
CC variant.  
XX  
SQ Sequence 197 AA:  
Query Match 96.5%; Score 980; DB 21; Length 197;  
Best Local Similarity 96.4%; Pred. No. 1.3e-94;  
Matches 190; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEVQIVSTATQTFLA 60  
DB 1 MKKGSVIVGRINLSGDTAYAQOTRGEQGTOKTSHTGRDKNQVGEVQIVSTATQTFLA 60  
QY 61 TSINGVLWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120  
DB 61 TSINGVLWTVYHGAGTTRTASPKGPVTOMYTNVDKDLVGMQAPQGSRSLLTPCTCGSSDLY 120  
QY 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGPGLPCPAGHAGVIFRAAVSTRGVAK 180  
DB 121 LVTRHADVIPVRRGDSRGSLLSPRISYLKSGSGPGLPCPAGHAGVIFRAAVSTRGVAK 180  
QY 181 AVDFIPVESLETTMRSP 197  
DB 181 AVDFIPVESLETTMRSP 197  
RESULT 5  
AAB15221  
ID AAB15221 standard; protein; 197 AA.  
XX  
AC AAB15221;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A-NS3 fusion protease #3.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX

DR WPI: 2000-465976/40.  
 DR N-PSDB; AAA73330.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Claim 23; Fig 13; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 94.8%; Score 963; DB 21; Length 197;  
 Best Local Similarity 94.9%; Pred. No. 7. Be-93;  
 Matches 187; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAOQTRGEGCTQKTSHTGRDNKNOVEGEVQIVSTATQTEFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAOQTRGEGCGEQTSQTRGRDNKNOVEGEVQIVSTAAQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTNVDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120  
 DB 61 TCINGVCTVYHGAGTRTIAAPKGPVTQMTNVDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGLLSPPRISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGLLSPPRISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 6  
 AAB15226  
 ID AAB15226 standard; protein; 197 AA.  
 XX  
 AC AAB15226;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #8.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;

XX  
 DR WPI: 2000-465976/40.  
 DR N-PSDB; AAA73335.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Example 5; Fig 18; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0  
 CC wild-type sequence.  
 XX  
 SQ Sequence 197 AA:  
 Query Match 92.1%; Score 936; DB 21; Length 197;  
 Best Local Similarity 93.9%; Pred. No. 5.4e-90;  
 Matches 185; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKKGSVVIVGRINLSGDTAYAOQTRGEGCTQKTSHTGRDNKNOVEGEVQIVSTATQTEFLA 60  
 DB 1 MKKGSVVIVGRINLSGDTAYAOQTRGLLCITSLTGRDNKNOVEGEVQIVSTAAQTFLA 60  
 QY 61 TSINGVLWTVYHGAGTRTIAAPKGPVTQMTNVDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120  
 DB 61 TCINGVCTVYHGAGTRTIAAPKGPVTQMTNVDKDLVGMQAPQGSRSLSLTPCTCGSSDLY 120  
 QY 121 LVTRHADVIPVRRGDSRGLLSPPRISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 DB 121 LVTRHADVIPVRRGDSRGLLSPPRISYLYKSGSGGPLLCPCPAGHAGVIFRAAVSTRGVAK 180  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 181 AVDFIPVESLETTMRSP 197  
 RESULT 7  
 AAB15220  
 ID AAB15220 standard; protein; 195 AA.  
 XX  
 AC AAB15220;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB: AAY73329.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX  
 PS Claim 23; Fig 12; 66pp; English.

XX The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease-inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.

SQ Sequence 195 AA;  
 Query Match 91.4%; Score 929; DB 21; Length 195;  
 Best Local Similarity 92.9%; Pred. No. 2.9e-89;  
 Matches 183; Conservative 3; Mismatches 9; Indels 2; Gaps 1;  
 QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEQGTOKTSHTGRDKNOVEGEVOIVSTATOTFLA 60  
 DB 1 MKKKGSVVIVGRIVLNG--AYAAQOTRGEQGTOKTSHTGRDKNOVEGEVOIVSTATOTFLA 58  
 QY 61 TSINGVLWTVYHGACGTRTIAASPKGPVTOMYTNVDKDLVGHQAPQGSRSLSLTPTCTGSSDLY 120  
 DB 59 TCINGVCWTVYHGACGTRTIAASPKGPVTOMYTNVDKDLVGHQAPQGSRSLSLTPTCTGSSDLY 118  
 QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 8  
 AAB15212  
 ID AAB15212 standard; protein: 195 AA.  
 XX  
 AC AAB15212;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO2000040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI: 2000-465976/40.  
 DR N-PSDB: AAY73328.  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 XX  
 PS Example 2; Fig 10; 66pp; English.

XX The present sequence is a fusion protein created using the Hepatitis C  
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
 CC essential for the replication of the virus, acting to cleave its  
 CC replicative proteins from the polyprotein produced from the HCV genome.  
 CC Inhibitors of the two proteins should be effective as antiviral  
 CC treatments of HCV infection. This is useful as HCV can lead to chronic  
 CC liver disease such as cirrhosis, liver failure and liver cancer. The  
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
 CC proteins which can be used to identify inhibitors of this type, as well  
 CC as enabling structural studies of the protease and protease-inhibitor  
 CC complexes.

SQ Sequence 195 AA;  
 Query Match 88.8%; Score 902; DB 21; Length 195;  
 Best Local Similarity 91.9%; Pred. No. 2e-86;  
 Matches 181; Conservative 1; Mismatches 13; Indels 2; Gaps 1;  
 QY 1 MKKKGSVVIVGRINLSGDTAYAAQOTRGEQGTOKTSHTGRDKNOVEGEVOIVSTATOTFLA 60  
 DB 1 MKKKGSVVIVGRIVLNG--AYAAQOTRGLLCIIITSLTGRDKNOVEGEVOIVSTAQTFLA 58  
 QY 61 TSINGVLWTVYHGACGTRTIAASPKGPVTOMYTNVDKDLVGHQAPQGSRSLSLTPTCTGSSDLY 120  
 DB 59 TCINGVCWTVYHGACGTRTIAASPKGPVTOMYTNVDKDLVGHQAPQGSRSLSLTPTCTGSSDLY 118  
 QY 121 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVSTRGVAK 180  
 DB 119 LVTRHADVIPVRRRGRSGSLSPRPISYLKSGSGGPGLLCPAGHAGVIFRAAVSTRGVAK 178  
 QY 181 AVDFIPVESLETTMRSP 197  
 DB 179 AVDFIPVESLETTMRSP 195

RESULT 9  
 AAY24943  
 ID AAY24943 standard; protein: 665 AA.  
 XX  
 AC AAY24943;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex SEQ ID NO:14.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO9928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.



```

XX (SCHE ) SCHERING CORP.
PA Malcolm BA, Taremi SS, Weber PC, Yao N;
PI WPI; 1999-385385/32.
XX New hepatitis C virus covalent complexes
XX Claim 6; Page 90-92; 21lpp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;
Query Match 86.2%; Score 875.5; DB 20; Length 665;
Best Local Similarity 85.2%; Pred. No. 6.8e-83;
Matches 167; Conservative 15; Mismatches 11; Indels 3; Gaps 1;
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQGTOKTSHTGRDKNOVEGEVQIVSTATQTFLAT 61
DB 22 GSVVIVGRILLSGSGSITAYSOOTRGLGCKKTSITGRDKNOVEGEVQIVSTATQTFLAT 81
QY 62 SINGVLMTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPIQMTYTNVDQDLVGWQAPPGARSILFPTCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217
RESULT 10
AAY24947
ID AAY24947 standard; Protein; 665 AA.
XX AC AAY24947;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:18.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```

```

PA (SCHE ) SCHERING CORP.
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
XX WPI; 1999-385385/32.
XX New hepatitis C virus covalent complexes
XX Claim 6; Page 100-102; 21lpp; English.
XX The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence represents a specifically claimed example of the above
CC complex. The covalent NS4A-NS3 complexes are useful for structural
CC determination and determination of mode of binding of HCV inhibitors by
CC NMR spectroscopy. They can also be used for detecting inhibitors of the
CC protease activity, the helicase activity and the ATPase activity of NS3.
CC The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC the non-covalent protease-peptide complexes previously available.
XX SQ Sequence 665 AA;
Query Match 85.9%; Score 872.5; DB 20; Length 665;
Best Local Similarity 84.7%; Pred. No. 1.4e-82;
Matches 166; Conservative 16; Mismatches 11; Indels 3; Gaps 1;
QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQGTOKTSHTGRDKNOVEGEVQIVSTATQTFLAT 61
DB 22 GSVVIVGRILLSGSGSITAYSOOTRGLGCKKTSITGRDKNOVEGEVQIVSTATQTFLAT 81
QY 62 SINGVLMTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGWQAPGQSRSLTPCTCGSSDLYL 121
DB 82 CVNGVCWTVYHGAGSKTLAGPKGPIQMTYTNVDQDLVGWQAPPGARSILFPTCTCGSSDLYL 141
QY 122 VTRHADVIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 181
DB 142 VTRHADVIPVRRRGDSRGLSPRISYLGKSSGGPLLCPCAGHAGVIFRAAVSTRGVAKA 201
QY 182 VDFIPVESLETTMRSP 197
DB 202 VDFVPVESMETTMRSP 217
RESULT 11
AAY24942
ID AAY24942 standard; Protein; 665 AA.
XX AC AAY24942;
XX DT 07-SEP-1999 (first entry)
XX DE HCV NS4A-NS3 complex SEQ ID NO:13.
XX KW HCV; hepatitis C virus; single chain recombinant complex; linker;
XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
XX hydrophobic domain; covalent complex; detection; inhibitor.
XX OS Hepatitis C virus.
XX OS Synthetic.
XX PN WO9928482-A2.
XX PD 10-JUN-1999.
XX PF 24-NOV-1998; 98WO-US24528.
XX PR 28-JUL-1998; 98US-0094331.
XX PR 28-NOV-1997; 97US-0067315.
XX PA (SCHE ) SCHERING CORP.

```

XX Malcolin BA, Taremi SS, Weber PC, Yao N;  
 XX  
 XX  
 DR WPI; 1999-385385/32.  
 XX  
 XX  
 PT New hepatitis C virus covalent complexes  
 XX  
 PS Claim 6; Page 88-90; 2ilpp: English.

```

20  QZ      Sequence      665 AA:
Query Match      85.8%; Score 871.5; DB 20; Length 665;
Best Local Similarity 85.2%; pred. No. 1.8e-82;
Matches 167; Conservative 14; Mismatches 12; Indels 3; Gaps 1;

Qy      5  GSWVLVGRINLSGD--TAVAAQOTRGSGGTOKTSHTRGROKNQVEGVQIVSTATOTFLAT 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      22  GSWVLVGRILISGSGSITAYSQOTRGLGCIKISLTRGROKNQVEGVQIVSTATQSFLAT 81
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      62  SINGVLVWYVHGAGTRIIASPKGPVTOMYTNVTKDLVGWQAPQGSRLTPCTCGSSDLYL 121
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      82  CVNGVCVWYVHGAGSKTLAGPKGITOMYTNVDQDLVGWQAPPGARSLTPCTCGSSDLYL 141
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      122  VTRHADVIVPVRRGDSHCSSLSPRISYLGKSSGGPLLCPAGHAVGIFPRAAYSTRGVAKA 181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      142  VTRHADVIVPVRRGDSRGSLSPRPVSYLGKSSGGPLLCPFGHAVGIFPRAAYCTRGVAKA 201
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      182  VDFIPVESLETTMRSP 197
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      202  VDFVPEVESMETTMRSP 217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12	
AA17880	
ID	AA17880 standard; Protein; 216 AA.
XX	
AC	AA17880;
XX	
DT	07-SEP-1999 (first entry)
XX	
DE	HCV NS4A-NS3 complex SEQ ID NO:4.
XX	
KW	HCV; hepatitis C virus; single chain recombinant complex; linker;
KW	NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW	hydrophobic domain; covalent complex; detection; inhibitor.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
PN	WO9928482-A2.
XX	
PD	10-JUN-1999.
XX	
XX	24-NOV-1998; 98WO-US24528.
XX	
PR	28-JUL-1998; 98US-0094331.
PR	28-NOV-1997; 97US-0067315.
XX	
XX	(SCHE ) SCHERING CORP.
PA	
XX	

PI	Malcolm BA, Taremi SS, Weber PC, Yao N:
XX	
XX	WPI; 1999-385385/32.
XX	
XX	New hepatitis C virus covalent complexes
XX	
XX	Claim 6; Page 76-77; 21pp; English.
XX	
CC	The present invention describes a covalent hepatitis C virus (HCV)
CC	NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
CC	NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC	hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC	to the amino terminus of the HCV NS3 protease domain. The present
CC	sequence represents a specifically claimed example of the above
CC	complex. The covalent NS4A-NS3 complexes are useful for structural
CC	determination and determination of mode of binding of HCV inhibitors by
CC	NMR spectroscopy. They can also be used for detecting inhibitors of the
CC	protease activity. The helicase activity and the ATPase activity of NS3.
CC	The covalent NS4A-NS3 complexes are more soluble, stable and active than
CC	the non-covalent protease-peptide complexes previously available.
XX	
XX	Sequence 216 AA:
XX	
XX	Sequence 216 SQ

Query Match.	85.5%;	Score	868.5;	DB	20;	Length	216;
Best Local Similarity	85.1%;	Pred.	No. 7.6e-83;				
Matches	166;	Conservative	15;	Mismatches	11;	Indels	3;
						Gaps	1;
Qy	5	GSVVIVGRINLGD	---TAAQOTRGEQGTQKTSHTGRDKNOVEGEVQIVSTATQTOTFLAT	61			
Db							
Qy	22	GSVVIVGRILISGSGSITATNSQOTRGLLGCCKTSLTGRDKNOVEGEVQVYSTATQSFAT	81				
Db							
Qy	62	SINGVLVTWVYHGAGTRTIASPKGPVTOMYTNVDKDLVGMQAPQGSRLTPTCTCGSSDLYL	121				
Db							
Qy	82	CVNGVCVTWVYHGAGSKTLAGPKGPIQOMYTNVDQDLVGMQAPPQARSULTPTCTCGSSDLYL	141				
Db							
Qy	122	VTRHADVIPYRRRCDSRGSLSPRPISYLGKSGSGPLLCPAGHAVGIFRAAVSPRGVAKA	181				
Db							
Qy	142	VTRHADVIPYRRRCDSRGSLSPRPISYLGKSGSGPLLCPSGHAVGIFRAAVCTRGVAKA	201				
Db							
Qy	182	VDFIPVESLETTMRS	196				
Db							
Qy	202	VDFIPVESMETTMS	216				
Db							

RESULT	l3
AAAY24946	
ID	AAAY24946 standard; Protein; 665 AA.
XX	
AC	AAAY24946;
XX	
DT	07-SEP-1999 (first entry)
XX	
DE	HCV NS4A-NS3 complex SEQ ID NO:17.
XX	
KW	HCV; hepatitis C virus; single chain recombinant complex; linker;
KW	NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW	hydrophobic domain; covalent complex; detection; inhibitor.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
PN	WO9928482-A2.
XX	
PD	10-JUN-1999.
XX	
Pf	24-NOV-1998; 98WO-US24528.
XX	
PR	28-JUL-1998; 98US-0094331.
PR	28-NOV-1997; 97US-0067315.
XX	
PA	(SCHE ) SCHERING CORP.
PI	Malcolm BA, Taremi SS, Weber PC, Yao N:

(SCHE) / SCHEMING CONF.  
Malcolm BA, Taremi SS, Weber PC, Yao N;

(SCHE) / SCHEMING CONF.  
Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.  
XX DR  
XX  
XX  
PT New hepatitis C virus covalent complexes  
XX  
PS Claim 6; Page 97-99; 21pp; English.

DR 1999-385385/32.  
XX  
XX  
PT New hepatitis C virus covalent complexes  
XX  
PS Claim 6; Page 85-87; 21pp; English.





Db 1005 RRCREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGDKNOVEGEVQIVST 1064  
QY 54 ATOTFLATSLNGVLTWYTHGAGTRTIASPKGPVTOMYTNVDKDLVGMWAPQGSRSILTPCT 113  
Db 1065 AAOFTLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMWAPQGSRSILTPCT 1124  
QY 114 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 173  
Db 1125 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

## RESULT 2

S40770  
genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: S40770; PC1285  
R:Okamoto, H.  
Submitted to the EMBL Data Library, March 1992  
A:Reference number: S40770  
A:Accession: S40770  
A:Molecule type: genomic RNA  
A:Residues: 1-3011 <OKA>  
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A:Reference number: PC1284; NID:91013116; PMID:2170712  
A:Accession: PC1285  
A:Molecule type: genomic RNA  
A:Residues: 1-513 <OK2>  
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512  
A:Experimental source: Isolate HC-J1  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:116-191/Product: capsid protein C #status predicted <CP>  
F:192-389/Product: major envelope protein E #status predicted <EPM>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 83.0%; Score 843.5; DB 1; Length 3011;  
Best Local Similarity 82.4%; Pred. No. 5e-68;  
Matches 168; Conservative 8; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAAQOTRGEQGTOKTSHTGRDKNOVEGEVQIVST 53  
Db 1005 RRCREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGDKNOVEGEVQIVST 1064  
QY 54 ATOTFLATSLNGVLTWYTHGAGTRTIASPKGPVTOMYTNVDKDLVGMWAPQGSRSILTPCT 113  
Db 1065 AAOFTLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMWAPQGSRSILTPCT 1124  
QY 114 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 173  
Db 1125 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

## RESULT 3

GNWVCH  
genome polyprotein - hepatitis C virus (strain H)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A36814; A41546  
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
submitted to Genbank, July 1992  
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus  
A:Reference number: A36814  
A:Accession: A36814  
A:Molecule type: genomic RNA  
A:Residues: 1-3011 <INC>  
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738  
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp  
A:Reference number: A41546; MUID:92052256; PMID:1658800  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F:116-191/Product: capsid protein C #status predicted <CP>  
F:192-389/Product: major envelope protein E #status predicted <EPM>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

## Query Match

Best Local Similarity 81.4%; Score 838.5; DB 1; Length 3011;  
Matches 166; Conservative 10; Mismatches 19; Indels 9; Gaps 1;

QY 3 KKGSVIVGRIN-----LSGDTAYAAQOTRGEQGTOKTSHTGRDKNOVEGEVQIVST 53  
Db 1005 RRCREILLGPADGMVSKGWRLLAPITAYAAQOTRGLGCIITSLTGDKNOVEGEVQIVST 1064  
QY 54 ATOTFLATSLNGVLTWYTHGAGTRTIASPKGPVTOMYTNVDKDLVGMWAPQGSRSILTPCT 113  
Db 1065 AAOFTLATCINGVCWTVYHGAGTRTIASPKGPVTOMYTNVDQDLVGMWAPQGSRSILTPCT 1124  
QY 114 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 173  
Db 1125 CGSSDLYLTVTRHADVTPVRRRGDSRGLSPRISYLGKSSGGPILCPAGHAGVGFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

## RESULT 4

GNWVW  
genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (strain Taiwan)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40244  
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 188, 102-113, 1992

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>		Query Match		80.5%; Score 817.5; DB 1; Length 3010;	
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>		Best Local Similarity		76.5%; Pred. No. 1.2e-65;	
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>		Matches		156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;	
QY	3	KKGSVVIVGRIN-----LSGDTAYAAQTGRGCGTKTSHTGRDNKNGVEGEVQIVST	53		
Db	1005	RRGREILLGPADSIEGQGWELLAPITAYAAQTGRLLGCIIVTSLTGRDNKNGVEGEVQIVST	1064		
QY	54	ATQTFATLSINGLVTVYHGACTRTIASPGPVTQMTYNVDKDLVGWQAPGGSRLTPCT	113		
Db	1065	ATQSFATLVGVCMTVEHGASKTLAPKGPITQMTYNVDODLVGWPDPGASRLTPCT	1124		
QY	114	CGSSDLYLVTRHADYIPVRRRGRDGRGSLSPRPISYLKSSGGPGLPCGHAVGIFRAAV	173		
Db	1125	CGSSDLYLVTRHADYIPVRRRGRDGRGSLSPRPVSYLKGSSGGPGLPCSHAVGIFRAAV	1184		
QY	174	STRGVAKAVDFIPVESLETHRSP	197		
Db	1185	CTRGVAKAVDFIPVESMETTRSP	1208		
<p>RESULT 6</p> <p>GNWVTC</p> <p>genome polyprotein - hepatitis C virus</p> <p>N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome isolated from h</p> <p>protein NS4a; nonstructural protein NS4b; nonstructural protein NS5</p> <p>C:Species: hepatitis C virus</p> <p>C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001</p> <p>C:Accession: A38465</p> <p>F:Takanizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, J. Virol. 65, 1105-1113, 1991</p> <p>A:Title: Structure and organization of the hepatitis C virus genome isolated from h</p> <p>A:Reference number: A38465; MUID:91140698; PMID:1847440</p> <p>A:Accession: A38465</p> <p>A:Molecule type: genomic RNA</p> <p>A:Residues: 1-3010 &lt;TAK&gt;</p> <p>A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAV72945.1; PID:g329771</p> <p>C:Superfamily: hepatitis C virus genome polyprotein</p> <p>C:Keywords: ARP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru</p> <p>F:2-115/Product: capsid protein C #status predicted &lt;CPC&gt;</p> <p>F:116-191/Product: envelope protein M #status predicted &lt;EPM&gt;</p> <p>F:192-389/Product: major envelope protein E #status predicted &lt;MEE&gt;</p> <p>F:390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;</p> <p>F:730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;</p> <p>F:1007-1615/Product: nonstructural protein NS3 #status predicted &lt;NS3&gt;</p> <p>F:1230-1237/Region: nucleotide-binding motif A (P-loop)</p> <p>F:1312-1317/Region: nucleotide-binding motif B</p> <p>F:1316-1319/Region: DEXH motif</p> <p>F:1616-1862/Product: nonstructural protein NS4a #status predicted &lt;N4A&gt;</p> <p>F:1863-2013/Product: nonstructural protein NS4b #status predicted &lt;N4B&gt;</p> <p>F:2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;</p> <p>F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2070</p>					
Query Match		80.1%; Score 813.5; DB 1; Length 3010;			
Best Local Similarity		76.0%; Pred. No. 2.7e-65;			
Matches		155; Conservative 21; Mismatches 19; Indels 9; Gaps 1;			
QY	3	KKGSVVIVGRIN-----LSGDTAYAAQTGRGCGTKTSHTGRDNKNGVEGEVQIVST	53		
Db	1005	RRGREILLGPADSIEGQGWELLAPITAYAAQTGRLLGCIIVTSLTGRDNKNGVEGEVQIVST	1064		
QY	54	ATQTFATLSINGLVTVYHGACTRTIASPGPVTQMTYNVDKDLVGWQAPGGSRLTPCT	113		
Db	1065	ATQSFATLVGVCMTVEHGASKTLAPKGPITQMTYNVDODLVGWPDPGASRLTPCT	1124		
QY	114	CGSSDLYLVTRHADYIPVRRRGRDGRGSLSPRPISYLKSSGGPGLPCGHAVGIFRAAV	173		
Db	1125	CGSSDLYLVTRHADYIPVRRRGRDGRGSLSPRPVSYLKGSSGGPGLPCPGHAVGIFRAAV	1184		





A:Accession: JC5620  
A:Molecule type: mRNA  
A:Residues: 1-3014 <CHA>  
A:Cross-references: GB:Y13184  
A:Experimental source: genotype 5a, which predominates in South Africa  
A:Note: the translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:384-408/Region: hypervariable #status predicted  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif A (P-loop)  
F:1317-1320/Region: DEXH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 73.2%; Score 743.5; DB 1; Length 3014;  
Best Local Similarity 68.8%; Pred. No. 6.4e-59;  
Matches 140; Conservative 25; Mismatches 30; Indels 9; Gaps 1;

QY 3 KRGSVVIVGRIN-----LSGDTAYAQQRGEGTQKTSHTGRDKNQVEQIVST 53  
DB 1006 RRGREIFLGPAADIKTSGRNLLAPITAYAQQRGVLGALVLSGTGRNREAGEVQFIST 1065  
QY 54 ATQTFLATISNGVLTVYHAGTRTASPKGPVTOMYTNVDKDLVGMQAPGSRSLTPTCT 113  
DB 1066 ATQTFLGICINGVMVTLFGAGSKTLAGPKGPVQMYTNVDKDLVGMQAPGSKSLTRCT 1125  
QY 114 CGSSDLYLVTRADYIPVRRGDSRGLSPRISVLKSGSGGPLLCPAGHAGVGFRAAV 173  
DB 1126 CGSADLYLVTRADYIPARRGDTASLLSPRISVLKSGSGGPLLCPGSHGVGFRAAV 1185  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
DB 1186 CTRGVAKALEFVPVENLETTMRSP 1209

RESULT 10  
JQ1303  
genome polyprotein - hepatitis C virus (isolate HC-J6)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C:Accession: JQ1303  
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.  
J. Gen. Virol. 72, 2697-2704, 1991  
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
A:Reference number: JQ1303; MUID:92044440; PMID:1658196  
A:Accession: JQ1303  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651  
A:Experimental source: isolate HC-J6 from a Japanese individual  
A:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1316-1323/Region: nucleotide-binding motif B  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

Query Match 66.8%; Score 679; DB 1; Length 3033;  
Best Local Similarity 69.3%; Pred. No. 4.7e-53;  
Matches 124; Conservative 27; Mismatches 28; Indels 0; Gaps 0;

QY 19 TAYAQQRGEGTQKTSHTGRDKNQVEQIVSTATQTFLATISNGVLTVYHAGTRT 78  
DB 1034 TAYAQQRGGLGIIVSVMTGRDTEQAGEIQVLSVTQSLGTTISGVLTVYHAGNKT 1093  
QY 79 IASPKGPVTOMYTNVDKDLVGMQAPGSRSLTCTCGSSDLYLVTRHADYIPVRRGDSR 138  
DB 1094 LAGSRGPVTOMYSSAEGDLVGMPSPPGTSLEPCTCGAVDLYLVTRNADYIPARRRDKR 1153  
QY 139 GSLLSPRISVLKSGSGGPLLCPAGHAGVGFRAAVSTRGVAKAVDFIPVESLETTMRSP 197  
DB 1154 GALLSPRLSTLKGSSGPPVLCPRGHAGVGFRAAVSRGVAKSIDFIPVETLIDIVTRSP 1212

RESULT 11  
GNVWJ8

genome polyprotein - hepatitis C virus (strain HC-J8)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C:Accession: A40250; PM0397; PM0559  
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda  
Virology 188, 331-341, 1992  
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to  
A:Reference number: A40250; MUID:92230232; PMID:1314459  
A:Accession: A40250  
A:Molecule type: genomic RNA  
A:Residues: 1-3033 <OKA>  
A:Cross-references: GB:D01098; GB:D01221; NID:g221608; PIDN:BAA01761.1; PID:g221609  
R:Chan, S.W.; Womish, F.; Holmes, E.C.; Dow, B.; Peuchere, J.F.; Follett, E.; Yap  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship  
A:Reference number: PM0393; MUID:92268871; PMID:1316939  
A:Accession: PM0397  
A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D01034  
A:Experimental source: isolate E-bl2  
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimo  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan  
A:Reference number: PM0554; MUID:92068204; PMID:1720309  
A:Accession: PM0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru-  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>  
F:1234-1241/Region: nucleotide-binding motif A (P-loop)  
F:1316-1321/Region: nucleotide-binding motif B  
F:1320-1323/Region: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20

Query Match 66.4%; Score 675; DB 1; Length 3033;  
Best Local Similarity 69.8%; Pred. No. 1.1e-52;  
Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;



A: Molecule type: DNA  
A: Residues: 1-590 <PAR>  
A: Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84658.1; PID:g738007  
A: Experimental source: serogroup A, strain 22491  
C: Genetics:  
A: Gene: NMA1418  
C: Superfamily: nitrate/nitrite sensor protein narX  
C: Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase; s  
F:395/Active site: His (phosphohistidine intermediate) #status predicted

	Query Match	8.8%	Score 89.5	DB 2	Length 590
	Best Local Similarity	20.8%	Pred. No. 2.9	Indels 65	Gaps 6
	Matches 45	Conservative 27	Mismatches 79		
QY	28	EQGTOKTSHTGRDKNQVEGVQIVSTATQTFLATSIINGVLWTVYHGAGTRTIASPKGPVT	87		
DB	213	EGGTPEFKQVGRCFNOMGGRKILYDDLEGOVALQ-----TRSLKQONOLT	259		
QY	88	OMYTNVDKDLVQWAPQ-----GSRSLPCTCGSSDLYLVTRHAD-----	127		
DB	260	LLY-QTTRDLHQSYIPOQAAEHFLNRLIPAYGADSGRVCLDGGSDVYVSIHHADCGTAAS	318		
QY	128	-----VIPVRRGDSRGSLSPRITSYLKSGSGPLICPAGHAVGIFRAAVSTR---	176		
DB	319	DLGKYHEEIFFIETQNETLGRLLLSFPNGISLDEDDRILLQTLGRQLGVSLAGAKQEEER	378		
QY	177	-----GVAKAYDF--IPVESLET	192		
DB	379	RLLAVLQERNLIAQGLHDSIAQALTFLNLQVOMLET	414		

Search completed: August 30, 2003, 19:20:32  
Job time : 17.2134 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 9.75674 Seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVVIVGRINLSGDTA.....YAKAVDIPVESLETHMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	844.5	83.1	3011	1 POLG_HCV1	P26664 h genome po
2	838.5	82.5	3011	1 POLG_HCVH	P27958 h genome po
3	826.5	81.3	3010	1 POLG_HCVTW	P29846 h genome po
4	817.5	80.5	3010	1 POLG_HCVJT	Q00269 h genome po
5	813.5	80.1	3010	1 POLG_HCVBK	P26663 h genome po
6	813.5	80.1	3010	1 POLG_HCVJA	P26662 h genome po
7	679	66.8	3033	1 POLG_HCVJG	P26660 h genome po
8	675	66.4	3033	1 POLG_HCVJ8	P26661 h genome po
9	90	8.9	321	1 HHOA_ARATH	Q9sel7 arabidopsis
10	85.5	8.4	209	1 PAD_PSEAE	O9bx08 pseudomonas
11	83.5	8.2	437	1 DEGL_ARATH	Q22609 arabidopsis
12	83	8.2	452	1 AMP_HUMAN	Q13685 homo sapien
13	78.5	7.7	485	1 Y136_TREPA	O83172 treponema p
14	78.5	7.7	660	1 VST2_HEVBU	P29326 hepatitis e
15	78.5	7.7	660	1 VST2_HEVPA	P33426 hepatitis e
16	78	7.7	401	1 FXH1_MOUSE	O88621 mus musculu
17	77.5	7.6	263	1 GBAK_MOUSE	O35205 mus musculu
18	77.5	7.6	301	1 MCP_BPF41	P26596 lactococcus
19	77.5	7.6	452	1 MLTD_ECOLI	P23931 escherichia
20	76.5	7.5	323	1 YPRT_SMRHV	P21407 squirrel mo
21	76.5	7.5	333	1 MOSA_RHIME	Q07607 rhizobium m
22	76	7.5	300	1 ERA_MYCLE	Q49768 mycobacteri
23	76	7.5	911	1 TB11_NEIMB	Q09056 neisseria m
24	76	7.5	3411	1 POLG_YEFV1	P03314 y genome po
25	76	7.5	3411	1 POLG_YEFV2	P19901 y genome po
26	76	7.5	3414	1 POLG_TBEVW	P14336 t genome po
27	75.5	7.4	248	1 TRY1_CHICK	P14336 t genome po
28	75.5	7.4	1425	1 NP4A_MOUSE	P59240 mus musculu
29	75.5	7.4	2269	1 WDR9_HUMAN	Q9nsi6 homo sapien
30	75	7.4	264	1 CTRL_HUMAN	P40313 homo sapien
31	75	7.4	300	1 SIAL_PIG	P31936 sus scrofa
32	75	7.4	467	1 NX1B_BOVIN	Q28142 bos taurus
33	74.5	7.3	248	1 GRAD_MOUSE	P11033 mus musculu

ALIGNMENTS

RESULT 1

POLG\_HCV1 STANDARD; PRT; 3011 AA.

AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE Hepatitis C virus (isolate 1) (HCV).

OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

ON NCBI\_TaxID=11104;

RX MEDLINE=91172826; PubMed=1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Ruo G., Houghton M.;

RA "Genetic organization and diversity of the hepatitis C virus.,"

Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

CC precursor polyprotein, commonly with Asp or Glu in the P6

CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

CC (RNA)(N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND MNNA.

CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

CC EMBL; M62321; AAA45676.1; --

CC PIR; A39166; GNMVC3.

CC PDB; 1ALV; 16-FEB-99.

CC PDB; 1HEI; 25-NOV-98.

CC MEROPS; S29.001; --

CC InterPro; IPR001410; DEAD.

CC InterPro; IPR002522; HCV\_capsid.

DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RDRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXdc; 1.  
 KW Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein: Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane: Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 192 381 MATRIX PROTEIN (POTENTIAL).  
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRG64;  
 Query Match 83.1%; Score 844.5; DB 1: Length 3011;  
 Best Local Similarity 82.4%; Pred. No. 3,4e-70;  
 Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;  
 3 KKGSVVIVGRIN-----LSGDYAYAOQTRGCGTQKTSHTGRKNQVGEVQIVST 53

Db 1005 RRGREILLGPADGMVSKGWRLAPITAYAAQOTRGLLGCIITSLTGDKKNQVGEVQIVST 1064  
 Qy 54 ATQTFATISNGVLTWYVYHGAGTRTITASPKGPTQMYTNVDKDLGVQWQAPQSGRSLSPTCT 113  
 Db 1065 AAQTFATISNGVLTWYVYHGAGTRTITASPKGPTQMYTNVDKDLGVQWQAPQSGRSLSPTCT 1124  
 Qy 114 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLVKLGSSGGLLCPAGHAGVGFRAAV 173  
 Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSPRISYLVKLGSSGGLLCPAGHAGVGFRAAV 1184  
 Qy 174 STCGVAKAVDFIPVESLETTMRSP 197  
 Db 1185 CTGCVAKAVDFIPVENLETTMRSP 1208  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchausti G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC [1-] FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC [1-] FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC [1-] FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC [1-] FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC [1-] FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC [1-] CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC [1-] CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC (RNA)(N).  
 CC [1-] SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M67463; AAA45534.1; .  
 DR PIR: A36814; GNMVCH.  
 DR PDB: 1HEI; 25-NOV-98.  
 DR PDB: 1A1V; 16-FEB-99.  
 DR PDB: 1A1R; 17-JUN-98.  
 DR MEROPS: S29.001; .  
 DR MEROPS: U39.001; .  
 DR TRANSFAC: T04155; .  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR001650; Helicase-C.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_Ps.  
 DR InterPro: IPR007094; RNA\_pol\_Psvir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase-C; 1.  
 DR Pfam: PF00998; Viral\_RdRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXdc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CHAIN 1 191 CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 192 383 CAPSID PROTEIN C.  
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.  
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.  
 FT CHAIN 810 1026 PROTEIN P7.  
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.  
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.  
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5A.  
 FT CHAIN 347 369 NONSTRUCTURAL PROTEIN NS5B.  
 FT ACT\_SITE 1083 1083 TRANSMEM.  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CHAIN 305 305 CARBOHYD  
 FT CHAIN 417 417 CARBOHYD  
 FT CHAIN 423 423 CARBOHYD  
 FT CHAIN 430 430 CARBOHYD  
 FT CHAIN 448 448 CARBOHYD  
 FT CHAIN 476 476 CARBOHYD  
 FT CHAIN 532 532 CARBOHYD  
 FT CHAIN 540 540 CARBOHYD  
 FT CHAIN 556 556 CARBOHYD  
 FT CHAIN 576 576 CARBOHYD  
 FT CHAIN 623 623 CARBOHYD  
 FT CHAIN 645 645 CARBOHYD  
 FT STRAND 1224 1226  
 FT TURN 1232 1233  
 FT TURN 1236 1238  
 FT HELIX 1239 1246  
 FT TURN 1247 1248  
 FT STRAND 1251 1255  
 FT HELIX 1258 1271  
 FT TURN 1272 1272  
 FT STRAND 1277 1280  
 FT TURN 1281 1282  
 FT STRAND 1283 1285  
 FT STRAND 1291 1295  
 FT TURN 1296 1301  
 FT TURN 1302 1303  
 FT STRAND 1312 1316  
 FT TURN 1317 1319  
 FT HELIX 1323 1335  
 FT TURN 1336 1340  
 FT TURN 1343 1347  
 FT TURN 1352 1353  
 FT TURN 1360 1361  
 FT STRAND 1362 1366  
 FT STRAND 1368 1368  
 FT STRAND 1373 1375  
 FT TURN 1376 1377  
 FT STRAND 1378 1380  
 FT HELIX 1382 1385  
 FT STRAND 1389 1393  
 FT HELIX 1397 1409  
 FT TURN 1410 1411  
 FT STRAND 1414 1417  
 FT TURN 1419 1420  
 FT STRAND 1432 1436  
 FT TURN 1438 1439  
 FT STRAND 1450 1453  
 FT STRAND 1456 1463  
 FT STRAND 1471 1478  
 FT STRAND 1480 1480  
 FT HELIX 1481 1488  
 FT TURN 1489 1490  
 FT STRAND 1497 1501  
 FT STRAND 1507 1507  
 FT STRAND 1511 1511  
 FT HELIX 1514 1527  
 FT HELIX 1532 1544  
 FT STRAND 1550 1550  
 FT HELIX 1555 1564  
 FT HELIX 1570 1578  
 FT TURN 1579 1580  
 FT HELIX 1584 1597  
 FT TURN 1598 1598  
 FT HELIX 1606 1611  
 FT TURN 1614 1618  
 FT STRAND 1622 1623  
 FT STRAND 1627 1627  
 FT STRAND 1635 1636  
 FT HELIX 1640 1652  
 FT SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;  
 Query Match 82.5%; Score 838.5; DB 1; Length 3011;  
 Best Local Similarity 81.4%; Pred. No. 1.2e-69;

Matches 166; Conservative 10; Mismatches 19; Indels 9; Gaps 1;

QY 3 KGSWVIVGRIN-----LSQDTAYAQOTRGEQTKTSHTGDKKNQVEGEQIVST 53  
DB 1005 RRGQEIILGPADGMSVSGWRLLAPITAYAQOTRGLLCITTSITGRDKKNQVEGEQIVST 1064  
QY 54 ATQTFLATISINGLWTVYHGAGTRTIASPGKPTVOMYTNVDKDLGWQAPQSGRSSTPCT 113  
DB 1065 ATQTFLATINCVCWTVYHGAGTRTIASPGKPVLOITYTNVDQDLVGPAPQSGRSSTPCT 1124  
QY 114 CGSDDLVLVTHADVIPRRRGDSRGLSPRISYILKSGSGGLPCPAGHAVGIFRAAV 173  
DB 1125 CGSDDLVLVTHADVIPRRRGDSRGLSPRISYILKSGSGGLPCPAGHAVGIFRAAV 1184  
QY 174 STRGVAKAVDFIPVESLETTMRSP 197  
DB 1195 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3  
POLG.HCVTW  
ID POLG.HCVTW STANDARD; PRT: 3010 AA.  
AC P29846;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Taiwan) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=31645;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230206; PubMed=1314449;  
RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
RT \*The Taiwanese hepatitis C virus genome: sequence determination and  
RL mapping the 5' termini of viral genomic and antigenomic RNA.\*;  
RL Virology 188:102-113(1992).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC (RNA)(N).  
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND RNA.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M84754; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A40244; GNMVTV.  
DR PDB; 1N64; 25-FEB-03.  
DR PDB; 1NS3; 08-APR-98.  
DR MEROPS; S29.001; -;  
DR MEROPS; U39.001; -;  
DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR004109; HCV\_NS3.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVlr.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR ProDom; PD186062; HCV\_NS1; 1.  
DR SMART; SM00487; DEXDC; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
KW 3D-structure.  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 233 233  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2529 2529  
FT CARBOHYD 2788 2788  
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 81.3%; Score 826.5; DB 1; Length 3010;  
Best Local Similarity 77.9%; Pred. No. 1.6e-68;  
Matches 159; Conservative 18; Mismatches 18; Indels 9; Gaps 1;

```
OY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQGTQKTSHTGRDNKNOVEGEVQIVST 53
DB 1005 RRGREILGADSLGEGRWLLAPITAYAAQOTIRGLFGICITSLGRDNKNOVEGEVQIVST 1064
OY 54 ATQTFEATLSINGVLWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPGSGSLPCT 113
DB 1065 ATQSFELATCINGCVTWVYHGAGSKTLAGPKGPITOMYTNVDODLVGWHAPGAGSLPCT 1124
OY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSGSGGGLLCPAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSGSGGGLLCPGSHGVGIFRAAV 1184
OY 174 STRGVAKAVDFIPVSELTWNSP 197
DB 1185 CTRGVAKAVDFVPVSEMETWNSP 1208

RESULT 4
POLG_HCVJT STANDARD: PRT: 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A. PubMed-1318627;
RA Tanaka T., Kato M., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D11168; BAA01943.1; -.
CC DR PIR: A45573; A45573.
CC DR PDB: 1AIQ; 25-MAR-98.
CC DR PDB: LJXP; 14-JAN-98.
CC DR MEROPS: S29.001; -.
CC DR MEROPS: U39.001; -.
CC DR InterPro: IPR001410; DEAD.
OY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQGTQKTSHTGRDNKNOVEGEVQIVST 53
DB 1005 RRGREILGADSLGEGRWLLAPITAYAAQOTIRGLFGICITSLGRDNKNOVEGEVQIVST 1064
OY 54 ATQTFEATLSINGVLWTVYHGAGTRTIASPKGPVTOMYTNVDKDLVGVQAPGSGSLPCT 113
DB 1065 ATQSFELATCINGCVTWVYHGAGSKTLAGPKGPITOMYTNVDODLVGWHAPGAGSLPCT 1124
OY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSGSGGGLLCPAGHAVGIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLYKSGSGGGLLCPGSHGVGIFRAAV 1184
OY 174 STRGVAKAVDFIPVSELTWNSP 197
DB 1185 CTRGVAKAVDFVPVSEMETWNSP 1208

InterPro: IPR002522; HCV_capsid.
InterPro: IPR002521; HCV_core.
InterPro: IPR002519; HCV_env.
InterPro: IPR002531; HCV_NS1.
InterPro: IPR002518; HCV_NS2.
InterPro: IPR004109; HCV_NS3.
InterPro: IPR000745; HCV_NS4a.
InterPro: IPR001490; HCV_NS4b.
InterPro: IPR002868; HCV_NS5a.
InterPro: IPR002166; HCV_NS5a.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR007094; RNA_pol_PSVir.
Pfam: PF01543; HCV_capsid; 1.
Pfam: PF01542; HCV_core; 1.
Pfam: PF01539; HCV_env; 1.
Pfam: PF01560; HCV_NS1; 1.
Pfam: PF01538; HCV_NS2; 1.
Pfam: PF02907; HCV_NS3; 1.
Pfam: PF01006; HCV_NS4a; 1.
Pfam: PF01001; HCV_NS4b; 1.
Pfam: PF01506; HCV_NS5a; 1.
Pfam: PF00271; helicase_C; 1.
Pfam: PF00998; Viral_RdRP; 1.
ProDom: PD186062; HCV_NS1; 1.
SMART: SM00487; DEXDC; 1.
Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
3D-structure.
INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
CHAIN 347 369 POTENTIAL.
ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
NP_BIND 1230 1237 ATP (POTENTIAL).
SITE 1316 1319 DECH BOX.
CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
Query Match 80.5%; Score 817.5; DB 1; Length 3010;
Best Local Similarity 76.5%; Pred. No. 1.le-67;
Matches 156; Conservative 20; Mismatches 19; Indels 9; Gaps 1;
OY 3 KKGSVVIVGRIN-----LSGDTAYAAQOTRGEQGTQKTSHTGRDNKNOVEGEVQIVST 53
```



```

Db 1005 RRREILLPADIEGCGRLAPITAYAQORGLGCVITSLTRDKNOVEGEVQVST 1064
Qy 54 ATOTFLATSLNGVLTIVYHGAGTIRTIASPKGVTQMTYNVDKDLVGMQAPQGSRLTPT 113
Db 1065 ATOSFLATCVNGVCTVFGAGSKTLGPKPITQMTYNVDQDLVGMHAPPCCARSLTPT 1124
Qy 114 CGSSDLYLTVRHADVTPVRRGRDGRSLSPRISVYLGKSSGGPILCPAGHAGVIFRAAV 173
Db 1125 CGSSDLYLTVRHADVTPVRRGRDGRSLSPRISVYLGKSSGGPILCPAGHAGVIFRAAV 1184
Qy 174 TRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 5
POLG_HCVBK STANDARD; PRT: 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (PC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE-96235224; PubMed=8647104;
RA Borowski P., Helland M., Oehlmann K., Becker B., Kornetevy L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE-97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moomaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE-98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
```

```

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M58335; AAA72945.1; -
CC FIR: A38465; GMYTC.
CC PDB: 1AIQ; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC PDB: 1NS3; 08-APR-98.
CC PDB: 1C2P; 15-NOV-00.
CC PDB: 1CSJ; 08-NOV-99.
CC PDB: 1GX5; 09-APR-02.
CC PDB: 1GX6; 10-APR-02.
CC PDB: 1QUV; 26-JUN-00.
CC PDB: 80HM; 20-APR-99.
CC MEROPS: S29.001; -
CC
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RORP.
CC InterPro: IPR007095; RNA_pol_DS_PS.
CC InterPro: IPR007094; RNA_pol_PSVir.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; Viral_RORP; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXDC; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
CC 3D-structure. 1 1
CC INIT_MET 1 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 729
CC CHAIN 730 1006
CC CHAIN 1007 1615
CC CHAIN 1616 1862
CC CHAIN 1863 2013
CC CHAIN 2014 3010
CC CHAIN 3010 369
CC TRANSMEM 347 1083
CC ACT_SITE 1083 1107
CC ACT_SITE 1107 1165
CC ACT_SITE 1165 1237
CC NP_BIND 1230 1319
CC SITE 1316 1319
```

FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1059 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1075 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1093 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1095 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1149 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1175 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT HELIX 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT TURN 1203 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT STRAND 1204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; F8422D5ECCFDFD9C CRC64;

Query Match 80.18; Score 813.5; DB 1; Length 3010;  
Best Local Similarity 76.08; Pred. No. 2.7e-67;  
Matches 155; Conservative 21; Mismatches 19; Indels 9; Gaps 1;  
Qy 3 KKGWVIVGRIN-----LSGDTAYAQOTRGEQCTQKTSHTGRDKNOVEGEVQIVST 53  
Db 1005 RRGKEILLGADSLGRLAPITAYSQTRGLGCIITSLTGRDKNOVEGEVQIVST 1064  
Qy 54 ATQTFATSLNGVLWTVYHGAGRTTASPKGVPTQMTYNDKDLVGWQAPQOGSRSLTPCT 113  
Db 1065 ATQTFATSLNGVLWTVYHGAGRTTASPKGVPTQMTYNDKDLVGWQAPQOGSRSLTPCT 1124  
Qy 114 CGSSDLYLVTRHADVIVPRRGDSRGSLLSPRPISYILKSGSGGPLLCAGHANGVIFRAAV 173  
Db 1125 CGSSDLYLVTRHADVIVPRRGDSRGSLLSPRPISYILKSGSGGPLLCAGHANGVIFRAAV 1184  
Qy 174 STRGVAKAVDFIPVESLETTMRSP 197  
Db 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 6  
ID POLG\_HCVJA STANDARD; PRT; 3010 AA.  
AC P26662;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.22.-); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Japanese) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91088550; PubMed=2175903;  
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
RA Sugimura T., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from  
RT Japanese patients with non-A, non-B hepatitis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
RL [2]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,  
RA Ohkoshi S., Shimotohno K.;  
RT "Molecular structure of the Japanese hepatitis C viral genome.";  
RL FEBS Lett. 280:325-328(1991).  
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; D90208; BAAL4233.1;  
CC PIR; A39253; GNMVQJ.  
CC HSP; P26663; IJXP.  
CC MEROPS; S29.001; -.  
CC MEROPS; O39.001; -.  
CC DEAD.  
CC InterPro; IPR001410;  
CC InterPro; IPR002522; HCV\_capsid.  
CC InterPro; IPR002521; HCV\_core.  
CC InterPro; IPR002519; HCV\_env.  
CC InterPro; IPR002531; HCV\_NS1.  
CC InterPro; IPR002518; HCV\_NS2.  
CC InterPro; IPR004109; HCV\_NS3.  
CC InterPro; IPR000745; HCV\_NS4a.  
CC InterPro; IPR001490; HCV\_NS4b.  
CC InterPro; IPR002868; HCV\_NS5a.  
CC InterPro; IPR002166; HCV\_RdRp.

[illegible]

```

DB      1185  CTRGVAKAVDFIPVSMETTRSP 1208
|||||
RESULT 7
POLG_HCVJ6          STANDARD;          PRT;   3033  AA.
ID      AC      POLG_HCVJ6          STANDARD;          PRT;   3033  AA.
AC      P26660;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE      Envelope glycoprotein E1 (GP32) (GP33); Envelope glycoprotein E2
DE      (GP68) (NS1); protein P7; Nonstructural protein NS2 (P21)
DE      (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE      (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE      NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE      NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS      Hepatitis C virus (isolate HC-J6) (HCV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11113;
RN      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=92044440; PubMed=1658196;
RX      Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA      Machida A., Miyakawa Y., Mayumi M.;
RT      "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT      from a human carrier: comparison with reported isolates for conserved
RT      and divergent regions.";
RL      J. Gen. Virol. 72:1697-2704(1991).
CC      -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC      HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC      NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC      -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC      precursor polyprotein, commonly with Asp or Glu in the P6
CC      position, Cys or Thr in P1 and Ser or Ala in P1'.
CC      -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC      {RNA}(N).
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA.
CC      -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collabora-
CC      tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in
CC      modified and this statement is not removed. Usage by and for commer-
CC      entities requires a license agreement (See http://www.isb-sib.ch/annou-
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D00944; BAA00792.1; -
DR      PTR; J01303; J01303.
DR      HSP; P27958; IHEI.
DR      MEROPS; S29.001; -
DR      MEROPS; U39.001; -
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR002522; HCV_capsid.
DR      InterPro; IPR002521; HCV_core.
DR      InterPro; IPR002519; HCV_env.
DR      InterPro; IPR002531; HCV_ns1.
DR      InterPro; IPR002538; HCV_NS2.
DR      InterPro; IPR004109; HCV_NS3.
DR      InterPro; IPR000745; HCV_NS4a.
DR      InterPro; IPR001490; HCV_NS4b.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_RdRP.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PSVIR.
DR      Pfam; PF01543; HCV_capsid; 1.

```



```

DR ProDom: PD186062; HCV_NS1: 1.
DR SHART; SH00487; DEXdc; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 3087
FT ACT_SITE 1087 1097
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 66.48; Score 675; DB 1; Length 3033;
Best Local Similarity 69.84; Pred. No. 2.1e-54;
Matches 125; Conservative 24; Mismatches 30; Indels 0; Gaps 0;

QY 19 TAYAQOTRGEGTQKTSHTGRDNQKQVEGQIVSTATOTFLATSLVGVVYHGAGIRT 78
DB 1034 TAYTQOTRGLGAIIVSVLTGRDNQKQAGOVVLSVTVQIFLCTISGLVTVYHGAGNKT 1093

QY 79 IASPKGPVQYNTVNDKLVQWQAPQSSRLTPTCGSSDLYLVTRHADVPIVRRKGDNR 138
DB 1094 LAGPKGPVQYNTVNDKLVQWQAPQSSRLTPTCGSSDLYLVTRHADVPIVRRKGDNR 1153

QY 139 GLLSPRISLYLKSSGSPILCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETNRSP 197
DB 1154 GALLSPRLSTLKGSSGSPVLCRSHAGVGLFRAAVCARGVAKSIDFIPVESLDVATRTP 1212

RESULT 9
ID HQAAARATH STANDARD; PRT: 321 AA.
AC Q9SEL7; O49507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).
GN HQAA OR AT4G18370 OR F2BJ12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lensch M.H.A., Sokolenko A., Herrmann R.G.;
RT "Identification and characterization of the chloroplast HhoA protease,
a homolog to the bacterial periplasmic protease HhoA.";
FT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC MEDLINE=20083488; PubMed=10617198;
RX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoer W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Krels B., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettitt A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Farmann B., Granderath K., Dauner D., Herzl A.,
RA Neubmann S., Argizou A., Vitale D., Liguori R., Piravandi E.,
RA Massenblat S., Hillier R., Schmidt W., Lecharny A., Aubourg S.,
RA Schnabl S., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Ghibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria L., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Clout J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Splith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";
RN Nature 402:769-777(1999).
RP [3]
RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.
RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
RA Kieselbach T.;
RT "The chloroplast lumen from Arabidopsis thaliana.";
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
gene model prediction. AT4G18370 and AT4G18375 were originally
fused into a single gene.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF114386; AAF24060.1; -  
 EMBL: AL021710; CAA16717.1; ALT\_SEQ.  
 EMBL: AL161548; CAB78839.1; ALT\_SEQ.  
 MEROPS: S01.279; -  
 InterPro: IPR001940; Protease2C.  
 InterPro: IPR001254; Ser\_protease\_Try.  
 Pfam: PF00089; trypsin; 1.  
 PRINTS: PR00834; PROTEASES2C.  
 KW Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.  
 FT TRANSIT 1 26 CHLOROPLAST (POTENTIAL).  
 FT TRANSIT 27 71 THYLAKOID.  
 FT CHAIN 72 321 PROTEASE HHOA.  
 FT DOMAIN 77 87 POLY-GLU.  
 FT ACT\_SITE 145 145 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 186 186 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 264 264 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT CONFLICT 40 40 R -> G (IN REF. 1).  
 SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;

Query Match 8.9%; Score 90; DB 1; Length 321;  
 Best Local Similarity 23.8%; Pred. No. 0.43;  
 Matches 55; Conservative 31; Mismatches 89; Indels 56; Gaps 11;

QY 2 KKKGSVVIVGRINL-----SGDTAAQOTRGSGT-----QKTSHTGRDKNOVEGEVQIV 51  
 DB 95 KTSPSVVYIEAIELPKTSGGDILTDENGKIEGTGSGFVWDKLGHI-----VTNYHVIA 148  
 QY 52 STATOTFLATSLNGVLTYYHAGTRTIAIPKGPVTQMYTNVDKDLVGWQAPQGSRLTP 111  
 DB 149 KLATDQF---GLORCKVSLVDKGR--FSKEGKIVGL--DPDNDLAVLKIEGRELNP 201  
 QY 112 CTCGSSDLVYTRHADVIPVRRGRDSRG-----SLLSPRPISYLK----- 151  
 DB 202 VVLGTSNDRVGQSCFAI-----GNPYGYENTLIGVVGSLGRLPEIPSPNGKSISEAQTQ 256  
 QY 152 -----GSSGGPLCPAGHAGVIFRAAVSTR--GVAKAVDF-IPVESLETTM 194  
 DB 257 ADINSNGSGPLDSYGHITGVTATFTPKSGMSSGVNFAPIDVTVVRTV 307

RESULT 10  
 ID PAAD\_PSEAE STANDARD; PRT; 209 AA.  
 AC Q9HX08.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Probable aromatic acid decarboxylase (EC 4.1.1.-).  
 GN PA4019.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Miziochuch S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE / PHENYLACRYLIC ACID DECARBOXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AF004818; AAG07406.1; -  
 PIR: H83144; H83144.  
 DR InterPro: IPR003382; Flavoprotein.  
 DR Pfam: PF02441; Flavoprotein; 1.  
 KW Hypothetical protein; Lyase; Decarboxylase; Complete proteome.  
 SQ SEQUENCE 209 AA; 22367 MW; 01FD081CC495D3F6 CRC64;

Query Match 8.4%; Score 85.5; DB 1; Length 209;  
 Best Local Similarity 27.9%; Pred. No. 0.67;  
 Matches 51; Conservative 16; Mismatches 61; Indels 55; Gaps 11;

QY 43 QVEGEVQ-IVSTATOTFLATSLNGVL-----WIVYHAGTRTIAIPKGPVTQMT 91  
 DB 29 QEREVHFLISKAAQLVMATETVALPAKPAQMAFLTEYCGAAGQI-----RVFG 80  
 QY 92 NVDKDLVGWQAPQGSRLTP-----CTCGSSDL-----YLVTRHADVIPVRRGRDS 137  
 DB 81 QND-----WMAPASGSSAPNAMYICPSTGTLSAVATGACNNLIERAADVALKER---- 131  
 QY 138 RGSLLSPR--PIS-----YLGSSGGPLCPAGHAGVIFRAAVSTRGVAKAVDFIPVES 189  
 DB 132 RPLVLYPREAPFSIHLENKLSNLGAVILPA--APGFYH---QPSQVEDLVDFVARI 186  
 QY 190 LET 192  
 DB 187 LNT 189

RESULT 11  
 ID DEGLARATH STANDARD; PRT; 437 AA.  
 AC 022609; O9LK85;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protease De-like 1, chloroplast precursor (EC 3.4.21.-).  
 GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RX MEDLINE=98175982; PubMed=9507020;  
 RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;  
 RT "Identification and characterization of DegP, a serine protease associated with the luminal side of the thylakoid membrane.";  
 RL J. Biol. Chem. 273:7094-7098(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones.";  
 RL DNA Res. 7:217-221(2000).  
 RN [3]  
 RP SEQUENCE OF 104-118.  
 RC STRAIN=cv. Columbia;  
 RA Kieselbach T., Bystedt M., Schroeder W.P.;  
 RL Submitted (JUL-2000) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO.  
CC CAN DEGRADE BETA-CASIN.  
CC -1- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND  
CC O-PHENANTHROLINE.  
CC -1- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID  
CC MEMBRANE.  
CC -1- INDUCTION: By heat shock.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.  
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC -----  
DR EMBL; AF028842; AAC39436.1; ?  
DR EMBL; AF028842; AAC39436.1; ?  
DR EMBL; AP001371; BAB02539.1; ?  
DR EMBL; AP001371; BAB02539.1; JOINED.  
DR MEROPS: S01.279; ?  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; Ser-protease\_Try.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00595; PDZ; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 1.  
DR PROSITE; PS0106; PDZ; 1.  
KW Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.  
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).  
FT TRANSIT 2 ? THYLAKOID.  
FT CHAIN 104 437 PROTEASE DO-LIKE 1.  
FT DOMAIN 152 321 SERINE PROTEASE.  
FT DOMAIN 324 421 PDZ.  
FT ACT\_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 280 280 CHARGE RELAY SYSTEM (POTENTIAL).  
FT CONFLICT 12 23 HSPSPQLSNST -> SSTFLHSPSSHL (IN REF.  
FT CONFLICT 36 36 V -> I (IN REF. 2).  
FT CONFLICT 54 54 P -> S (IN REF. 2).  
FT CONFLICT 60 60 G -> R (IN REF. 2).  
FT CONFLICT 64 64 G -> D (IN REF. 2).  
FT CONFLICT 68 69 LL -> HF (IN REF. 2).  
FT CONFLICT 355 355 L -> V (IN REF. 2).  
FT CONFLICT 381 381 I -> V (IN REF. 2).  
FT CONFLICT 416 416 Q -> E (IN REF. 2).  
SQ SEQUENCE 437 AA; 48213 MW; 1497B1AB3F5FF2A4 CRC64;  
Query Match 8.2%; Score 83.5; DB 1; Length 437;  
Best Local Similarity 26.2%; Pred. No. 2.5;  
Matches 45; Conservative 17; Mismatches 55; Indels 55; Gaps 7;  
QY 70 VYHGAGTRTASPKGPVTQMY-----TNVDKDLVGV-----QA 102  
DB 150 VPQSGSGFVMDQGHIVTYNHVIRGASDLRVTLADQTTDFDAKVGVGDQDKVAVLRIDA 209  
QY 103 PGGRSLTPTCTGSSDLYLV-----TRHADVIPVRRRGDSRGLSPRI 147  
DB 210 PK-NKLRLPIPVGSADLLVGVKQVFAIGNPFGDLHTLTGIVISGLRRETS--SAATGRPI 265  
QY 148 SYL-----KGSSGGLPLCPAGHAVIPRAAVSTRGVAKAVDF-IPVESL 190  
DB 266 QDVITQDAINPGNSGGLDSSGTLIGINTAIYSPGASSGVGFSPIDTV 317  
RESULT 12  
AAMP\_HUMAN STANDARD; PRT; 452 AA.  
AC Q13685;  
DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Angio-associated migratory cell protein.  
GN AAMP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=95262124; PubMed=7743515;  
RA Beckner M.E., Krutzsch H.C., Stracke M.L., Williams S.T.,  
RA Gallardo J.A., Liotta L.A.;  
RT \*Identification of a new immunoglobulin superfamily protein expressed  
RT in blood vessels with a heparin-binding consensus sequence.\*;  
RL Cancer Res. 55:2140-2149(1995).  
CC -1- FUNCTION: MAY HAVE A FUNCTION IN MIGRATING CELLS.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BLOOD VESSELS. STRONGLY EXPRESSED  
CC IN ENDOTHELIAL CELLS, CYTOTROPHOBLASTS, AND POORLY DIFFERENTIATED  
CC COLON ADENOCARCINOMA CELLS FOUND IN LYMPHATICS.  
CC -1- SIMILARITY: Contains 8 WD repeats.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
CC -----  
DR EMBL; M95627; AAA68889.1; ?  
DR PIR; I39383; I39383.  
DR Genew; HGNC:18; AAMP.  
DR MIM; 603488; ?  
DR GO; GO:0008201; F:heparin binding activity; TAS.  
DR InterPro: IPR001880; WD40.  
DR Pfam: PF00400; WD40; 8.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
DR PROSITE; PS00682; WD\_REPEATS\_2; 6.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
DR Repeat; WD repeat.  
FT DOMAIN 14 18 HEPARIN-BINDING (POTENTIAL).  
FT DOMAIN 71 77 POLY-GLO.  
FT REPEAT 107 138 WD 1.  
FT REPEAT 150 180 WD 2.  
FT REPEAT 190 220 WD 3.  
FT REPEAT 231 261 WD 4.  
FT REPEAT 276 306 WD 5.  
FT REPEAT 333 363 WD 6.  
FT REPEAT 374 404 WD 7.  
FT REPEAT 416 446 WD 8.  
SQ SEQUENCE 452 AA; 49015 MW; DA1413D25EB236C0 CRC64;  
Query Match 8.2%; Score 83; DB 1; Length 452;  
Best Local Similarity 25.3%; Pred. No. 2.9;  
Matches 42; Conservative 13; Mismatches 47; Indels 64; Gaps 9;  
QY 68 WTVYHGAGTRTASPKGPVTQMYTNVDKDLVGVQAPQGRSL-----TPCTGSSDLYLV 122  
DB 197 WMEH-----PRAPVLLAGT-ADGNTWMMKVPNGDCKTFQGNPCATCGR----- 240  
QY 123 TRHADVIPVRRR---GDSRGS-----LLSPRPISYLGSSG--GPLLCPA----- 162  
DB 241 -----VLPDGRVAVGYEDGTIRINDLAKQSPFIRVLKGTGSHOGLTCVAAHQDGLILT 295  
QY 163 -----CHAVGIFR-----AAVSTRGVAKAVDFIPVESL 190  
DB 296 GSVDCQAKLVSAATTKGVGVFRPTVAVSOPSLGEGESESNSVESL 341  
RESULT 13



Y136_TREPA	STANDARD;	PRT; 485 AA.
ID	Y136_TREPA	
AC	O83172;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Hypothetical lipoprotein TP0136 precursor.	
GN	TP0136.	
OS	Treponema pallidum.	
OC	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.	
OX	NCBL_TaxID=160;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Nichols;	
RX	MEDLINE=98332770; PubMed=9665876;	
RA	Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,	
RA	Dodson R., Gwyn M., Hickey E.K., Clayton R., Ketchum K.A.,	
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,	
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,	
RA	McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,	
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,	
RA	Venter J.C.;	
RT	*Complete genome sequence of Treponema pallidum, the syphilis	
RT	spirochete.*	
RL	Science 281:375-388(1998).	
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor	
CC	(Potential).	
CC	-1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation	
CC	at the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/	
CC	or send an email to license@isb-sib.ch).	
CC		
DR	EMBL; AF001199; AAC65137.1; ALT_INIT.	
DR	TIGR; TP0136; -	
KW	Hypothetical protein; Lipoprotein; Membrane; Signal;	
KW	Complete proteome.	
FT	SIGNAL 1 23	POTENTIAL.
FT	CHAIN 24 485	HYPOTHETICAL LIPOPROTEIN TP0136.
FT	LIPID 24 24	N-ACYL-DIGLYCERIDE (POTENTIAL).
FT	DOMAIN 164 178	GLY/SER-RICH.
FT	DOMAIN 196 210	GLY/SER-RICH.
FT	DOMAIN 253 267	GLY/SER-RICH.
FT	DOMAIN 318 327	POLY-SER.
FT	DOMAIN 444 447	POLY-SER.
SQ	SEQUENCE 485 AA; 48984 MW; C7A4CEEDC7DC5CED CRC64;	
	Query Match 7.7%; Score 78.5; DB 1; Length 485;	
	Best Local Similarity 23.4%; Pred. No. 8.3;	
	Matches 50; Conservative 16; Mismatches 77; Indels 71; Gaps 10;	
Qy	16 SGDAYA-----QQTRGGGQTQKTH---- <td>63</td>	63
Db	54 AGSKLYATNRLWEKELNGTGSMOKVSSSVPTDSDK-----KVMSIATDGTNTFLVACVP	108
Qy	64 -NGVLWTIVYHCAG--TRTIASPKPYTMYNVDKDLVG-----NQAPQGSRLTPTCT	113
Db	109 GTGVYKHCVNGAGSSSTGTITASPTSETCSQHAT----LVGGTSKPFLLVPGGTGNNCGC	164
Qy	114 C-----GSSDLLYLVTIRADVIP-----VRRRGDSRGSLLSPRISYLK-----	151
Db	165 CGGGGGSSSSSSSSCIHWLPVPGTGNNGNCGCCGGGGSSSSSSSCIIHKVENTDEOFL	224
Qy	152 -----GSSGGPLLCPAGHAVG 167	
Db	225 DMGEGYVVTTKHLTKNGSSSAGPAQCPCGGGGG 258	

RESULT 14



```
RN SEQUENCE FROM N.A.
RX MEDLINE-92115700; PubMed-1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters I.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80581; AAA5727.1; -.
DR InterPro; IPR004261; SP2.
DR Pfam; PF03014; SP2; 1.
KW Signal.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB46FD3 CRC64;

Query Match 7.7%; Score 78.5; DB 1; Length 660;
Best Local Similarity 19.7%; Pred. No. 12;
Matches 45; Conservative 42; Mismatches 85; Indels 57; Gaps 10;

QY 10 VGRINLSGDTAYAAQTRGEQGTQKTSHTGTRDKNOV---EGEVQIVYSTATQTFLA---T 61
DB 297 LGULDFALELEFNLTGNTNTRVSRYSSTARHRLRGADGTAELTTTAATRFMKDLYFT 356

QY 62 SINGV-----LWTVYHGAGT-----RTIASPKG-PVTOMYTNV 93
DB 357 STNGVGEIGRGIALTLFNLADTLGLGLTELISSAGQLFYSRPVVSANGEPTVKLYTSV 416

QY 94 DKDLVGHQAQCGSRSLTPCTCGSSDLYLV---YRHADVIPVRRGDSRG-SILLSRPISY 149
DB 417 ENA----QODKGIAPHDIDLGESRVVIQDYDNOHEQDRTPSPAPSPFSLRANDVLW 472

QY 150 LK-----GSSGGPLLCAGHAGVIFRAAVSTRGVAKAVDFIPV 187
DB 473 LSLTAHYDOSTYCGSSGTPVY--VSDSVTLVNVATGQAQVARSIDWTKV 519
```

Search completed: August 30, 2003, 19:13:50  
Job time : 10.7567 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 37.5921 Seconds  
(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*
- 15: sp.virus:\*
- 16: sp.bacteriap:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862.5	84.9	4040	12 Q91FH8	Q91FH8 mucosal dis
2	848.5	83.5	3011	12 Q36579	Q36579 hepatitis c
3	844.5	83.1	2436	12 Q81756	Q81756 hepatitis c
4	844.5	83.1	3011	12 Q91FE5	Q91FE5 hepatitis c
5	844.5	83.1	3011	12 Q91ES8	Q91ES8 hepatitis c
6	843.5	83.0	3011	12 Q03463	Q03463 hepatitis c
7	841.5	82.8	3011	12 Q36608	Q36608 hepatitis c
8	841.5	82.8	3015	12 Q9PMX5	Q9PMX5 hepatitis c
9	841.5	82.8	3015	12 Q9PMW9	Q9PMW9 hepatitis c
10	839	82.6	181	12 Q91RR8	Q91RR8 hepatitis c
11	839	82.6	181	12 Q91R55	Q91R55 hepatitis c
12	837	82.4	181	12 Q91R55	Q91R55 hepatitis c
13	837	82.4	181	12 Q91RR2	Q91RR2 hepatitis c
14	837	82.4	181	12 Q91RT9	Q91RT9 hepatitis c
15	836	82.3	181	12 Q91RR3	Q91RR3 hepatitis c
16	836	82.3	181	12 Q91RR4	Q91RR4 hepatitis c

17	836	82.3	181	12 Q91RS1	Q91RS1 hepatitis c
18	836	82.3	181	12 Q91RQ8	Q91RQ8 hepatitis c
19	836	82.3	181	12 Q91RT1	Q91RT1 hepatitis c
20	836	82.3	181	12 Q91RR0	Q91RR0 hepatitis c
21	835.5	82.2	3011	12 Q36609	Q36609 hepatitis c
22	834	82.1	181	12 Q91RR6	Q91RR6 hepatitis c
23	834	82.1	181	12 Q91RS9	Q91RS9 hepatitis c
24	833	82.0	181	12 Q91RS3	Q91RS3 hepatitis c
25	832.5	81.9	3011	12 Q9DIT6	Q9DIT6 hepatitis c
26	832	81.9	181	12 Q91RT4	Q91RT4 hepatitis c
27	832	81.9	181	12 Q91RS8	Q91RS8 hepatitis c
28	832	81.9	181	12 Q91RT3	Q91RT3 hepatitis c
29	832	81.9	181	12 Q91RS5	Q91RS5 hepatitis c
30	832	81.9	181	12 Q91RS7	Q91RS7 hepatitis c
31	832	81.9	181	12 Q91RT0	Q91RT0 hepatitis c
32	832	81.9	181	12 Q91RS2	Q91RS2 hepatitis c
33	831	81.8	181	12 Q91RS6	Q91RS6 hepatitis c
34	830.5	81.7	3010	12 Q9QP61	Q9QP61 hepatitis c
35	830	81.7	181	12 Q91RS4	Q91RS4 hepatitis c
36	829.5	81.6	3010	12 Q68533	Q68533 hepatitis c
37	829	81.6	181	12 Q91RR7	Q91RR7 hepatitis c
38	829	81.6	181	12 Q91RT6	Q91RT6 hepatitis c
39	828	81.5	3011	12 Q36610	Q36610 hepatitis c
40	827.5	81.4	361	12 Q91RT8	Q91RT8 hepatitis c
41	827.5	81.4	361	12 Q70818	Q70818 hepatitis c
42	827.5	81.4	361	12 Q70817	Q70817 hepatitis c
43	827	81.4	181	12 Q91RR9	Q91RR9 hepatitis c
44	826.5	81.3	3010	12 Q9DTE2	Q9DTE2 hepatitis c
45	826.5	81.3	3010	12 Q99AU2	Q99AU2 hepatitis c

ALIGNMENTS

RESULT 1

ID	Q91FH8	PRELIMINARY;	PRT; 4040 AA.
AC	Q91FH8;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DR	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	Genome polyprotein.		
OS	Mucosal disease virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;		
OX	Pestivirus.		
NCBI_TaxID	11099;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20323484; PubMed-10864644;		
RA	Lai V.C., Zhong W., Skelton A., Ingravallo P., Vassilev V.,		
RA	Donis R.O., Hong Z., Lau J.Y.;		
RT	*Generation and characterization of a hepatitis C virus NS3 protease-		
RT	dependent bovine viral diarrhea virus.*;		
RL	J. Virol. 74:6339-6347(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Lai V.C.H., Hong Z.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF268278; AAF82566.1;		
DR	HSSP; P26663; LJP.		
DR	MEFOPS; S31.001;		
DR	InterPro; IPR000280; CDvir_endptsep80.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR001005; Myb_DNA_binding.		
DR	InterPro; IPR001568; RNase_T2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PS_vir.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		

```
DR PRINTS; PR00729; CDVENDOPTASE.
DR SMART; SM00487; DEXDC; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS05007; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
DR PROSITE; PS00531; RNASE_T2-2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase.
KW SEQUENCE 4040 AA; 453073 MW; ADE877910055B9DC CRC64;

Query Match      84.9%; Score 862.5; DB 12; Length 4040;
Best Local Similarity 88.7%; Pred. No. 1.9e-72;
Matches 173; Conservative 5; Mismatches 14; Indels 3; Gaps 1;

QY 5 GSVVIVGRINLSGD---TAYAQOTRGEQGTQKTSHTGRDNQVEGEVQIVSTATQFLAT 61
DB 10 GSVVIVGRINLSGSGSITACAAQTGRGLGCKITSLTGRDNQVEGEVQIVSTATQFLAT 69
QY 62 SINGLVTVYHGAGTRTIAAPKGPVQMTYNTVDKLDVGMQAPQGSRLTPTCTCGSSDLYL 121
DB 70 CINGVCVTVYHGAGTRTIAAPKGPVQMTYNTVDQDLVGMWPAQGSRLTPTCTCGSSDLYL 129
QY 122 VTRHADVIVPVRRGDSRGLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVSTRGVAKA 181
DB 130 VTRHANVIVPVRRGDSRGLSPRPISYLKSGSGGPLLCAGHAGVIFRAAVCTRGVAKA 189
QY 182 VDFIPVESLTTRS 196
DB 190 VDFIPVENLETTTRS 204

RESULT 2
Q36579 ID Q36579 PRELIMINARY; PRT; 3011 AA.
AC Q36579
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RL transcribed RNA.";
RL Science 277:570-574(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002530; HCV NS2.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR004109; HCV NS3.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR007094; RNA_pol_PSVir.
```

```
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01560; HCV_NS1; 1.
DR PFAM; PF01536; HCV_NS2; 1.
DR PFAM; PF02907; HCV_NS3; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF00271; Helicase_C; 1.
DR PFAM; PF00998; Viral_RDRP; 1.
DR PRODOM; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05021; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
KW SEQUENCE 3011 AA; 327182 MW; E2E0EE809C63C1B9 CRC64;

Query Match      83.5%; Score 848.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 2.8e-71;
Matches 168; Conservative 10; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGEQGTQKTSHTGRDNQVEGEVQIVST 53
DB 1005 RRGQILLGPGADGMVSKGWRLLAPITAYAQOTRGLGCIITSLTGRDNQVEGEVQIVST 1064
QY 54 ATOTFLATISINGLVTVYHGAGTRTIAAPKGPVQMTYNTVDKLDVGMQAPQGSRLTPTCT 113
DB 1065 ATQTFATCINGVCVTVYHGAGTRTIAAPKGPVQMTYNTVDQDLVGMWPAQGSRLTPTCT 1124
QY 114 CGSSDLXLVTRHADVIVPVRRGDSRGLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 173
DB 1125 CGSSDLXLVTRHADVIVPVRRGDSRGLSPRPISYLKSGSGGPLLCAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 3
Q81756 ID Q81756 PRELIMINARY; PRT; 2436 AA.
AC Q81756
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein (fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Choo Q.-L., Richman K., Han J.;
RT "The nucleotide sequence of the Hepatitis C viral genome.";
RL Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.
DR EMBL; M32084; AAA45677.1; -.
DR HSSP; P27938; 1AIV.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR PFAM; PF01560; HCV_NS1; 1.
DR PFAM; PF01538; HCV_NS2; 1.
```

```

DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS0507; RDRP_POSITIVE; 1.
DR PROSITE; PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
FT NON_TER 1
FT TER 2436
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 83.1%; Score 844.5; DB 12; Length 2436;
Best Local Similarity 82.4%; Pred. No. 5.1e-71;
Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEQGTOKTSHTGRDKNOVEGEVQIVST 53
DB 555 RGREILLGPADGVSKGWRLLAPITAYAAQTRGLLCITSLTGRDKNOVEGEVQIVST 614
QY 54 ATQTFLATLSINGLVLTYYHAGTRTIIASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCT 113
DB 615 AAQTFLATCINGVCWTYYHAGTRTIIASPKGPVIQMTYNDQDLVGWPAQGSRLTPTCT 674
QY 114 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 173
DB 675 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 734
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 735 CTRGVAKAVDFIPVENLETTMRSP 758

RESULT 4
ID Q9IF5 PRELIMINARY; PRT; 3011 AA.
AC Q9IF5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
CC -/- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF271632; AAF81759.1; -.
DR HSSP; P27958; 1AIV.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.

```

```

DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS0507; RDRP_POSITIVE; 1.
DR PROSITE; PS0521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327124 MW; 2489CE74AC864E58 CRC64;

Query Match 83.1%; Score 844.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 6.7e-71;
Matches 168; Conservative 9; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAAQTRGEQGTOKTSHTGRDKNOVEGEVQIVST 53
DB 1005 RGREILLGPADGVSKGWRLLAPITAYAAQTRGLLCITSLTGRDKNOVEGEVQIVST 1064
QY 54 ATQTFLATLSINGLVLTYYHAGTRTIIASPKGPVTQMTYNDKDLVGWQAPQGSRLTPTCT 113
DB 1065 AAQTFLATCINGVCWTYYHAGTRTIIASPKGPVIQMTYNDQDLVGWPAQGSRLTPTCT 1124
QY 114 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRRGDSRGLSPRPISYLKSGSGGPLLCPAGHAGVIFRAAV 1184
QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 5
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=colonel;
RA Desai S.M., Devare S., Yamaguchi J.;
RT "Hepatitis C Virus.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -/- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF290978; AAG02099.1; -.
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.

```

```

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327107 MW; A6BECF5A3B3EE13F CRC64;

Query Match
Best Local Similarity 81.1%; Score 844.5; DB 12; Length 3011.
Matches 167; Conservative 11; Mismatches 17; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDTAYAQOTRGQGTQKTSHTGRDKNQVEGEVQIVST 53
Db 1005 RRGQILLGPADGMVSKGWRLLAPITAYAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTPLATISNGVLVTVYHAGTRTIASPKGPVQMTYNTVDKLVGWAQPGSRSLTPCT 113
Db 1065 ATOTPLATISNGVLVTVYHAGTRTIASPKGPVQMTYNTVDQDLVGPAPGCSRLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGDSRGLSLSPRPISYLKGGSGGGLLCPAGHVGIFRAAV 173
Db 1125 CGSSDLYLVTRHADVIPVRRGDSRGLSLSPRPISYLKGGSGGGLLCPAGHVGIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1185 CTRGVAKAVDFIPVENLETTMRSP 1208

RESULT 6
Q03463 ID Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H.;

```

```

RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kural K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-J1;
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchausti G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255(1994).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D10749; BAA01582.1; -.
DR HSSP: P27958; 1HE1.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_Ps.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 97E9052C0250463B CRC64;

```

```
Query Match      83.0%; Score 843.5; DB 12; Length 3011;
Best Local Similarity 82.4%; Pred. No. 8.4e-71;
Matches 168; Conservative 8; Mismatches 19; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDYAYAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVST 53
DB 1005 RRGQEILLGPADGMVSKGWRLLAPITAYAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATSIINGVLTWYVHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPQGSRLTPCT 113
DB 1065 AAQTFLATCINGCVWTYVHGAGTRTIASPKGPVTOMYTNVDQDLVGMWPAQGSRLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 7
O36608 PRELIMINARY; PRT; 3011 AA.
AC O36608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77.
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee."
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF011751; AAB67036.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
```

```
DR Pfam: PF00998; Viral_RdRP; 1.
DR PRODOM: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327112 MW; 0B75E6B81CB5C198 CRC64;

Query Match      82.8%; Score 841.5; DB 12; Length 3011;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

QY 3 KGSVVIVGRIN-----LSGDYAYAQOTRGEQGTQKTSHTGRDKNQVEGEVQIVST 53
DB 1005 RRGQEILLGPADGMVSKGWRLLAPITAYAQOTRGLGCIITSLTGRDKNQVEGEVQIVST 1064

QY 54 ATOTFLATSIINGVLTWYVHGAGTRTIASPKGPVTOMYTNVDKLVGMQAPQGSRLTPCT 113
DB 1065 AAQTFLATCINGCVWTYVHGAGTRTIASPKGPVTOMYTNVDQDLVGMWPAQGSRLTPCT 1124

QY 114 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 173
DB 1125 CGSSDLYLVTRHADVIPVRRGRDGRGSLSPRPISYLGKSSGGPLLCPCAGHAGVIFRAAV 1184

QY 174 STRGVAKAVDFIPVESLETTMRSP 197
DB 1185 CTRGVAKAVDFIPVESLETTMRSP 1208

RESULT 8
O9PWX5 PRELIMINARY; PRT; 3015 AA.
AC O9PWX5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras."
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177040; AAF01182.1; -.
DR EMBL: AF177038; AAF01180.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
```

```

DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BC1F190663A CRC64;

Query Match 82.8%; Score 841.5; DB 12; Length 3015;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAOOTRGEQGTQKTSHTGRDKNQVEGEVIVST 53
Db 1009 RRGQELLGPADGMVSKGWRLLAPITAYAOOTRGLGCIITSITGRDKNQVEGEVIVST 1068

Qy 54 ATQTFATLSINGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 113
Db 1069 ATQTFATCINGVCTVYHGAGTRTIASPKGPVTQMTYNDQDLVGWPAQGSRLTPCT 1128

Qy 114 CGSSDLVLTVRHADVIPVRRGDSRGLSPRISYLKSGSGGGLLCPAGHVGIFRAAV 173
Db 1129 CGSSDLVLTVRHADVIPVRRGDSRGLSPRISYLKSGSGGGLLCPAGHVGIFRAAV 1188

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 9
Q9PM09 PRELIMINARY; PRT; 3015 AA.
AC Q9PM09;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.";
RL Virology 262:250-263(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Bukh J.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

```

```

CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF177039; AAF01181.1; -.
DR EMBL: AF177037; AAF01179.1; -.
DR HSSP: P27958; 1HEI.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR002129; Pyridoxal_deC.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc; 1.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 82.8%; Score 841.5; DB 12; Length 3015;
Best Local Similarity 81.9%; Pred. No. 1.3e-70;
Matches 167; Conservative 10; Mismatches 18; Indels 9; Gaps 1;

Qy 3 KGSVVIVGRIN-----LSGDTAYAOOTRGEQGTQKTSHTGRDKNQVEGEVIVST 53
Db 1009 RRGQELLGPADGMVSKGWRLLAPITAYAOOTRGLGCIITSITGRDKNQVEGEVIVST 1068

Qy 54 ATQTFATLSINGVLTWVYHGAGTRTIASPKGPVTQMTYNDKDLVGWQAPQGSRLTPCT 113
Db 1069 ATQTFATCINGVCTVYHGAGTRTIASPKGPVTQMTYNDQDLVGWPAQGSRLTPCT 1128

Qy 114 CGSSDLVLTVRHADVIPVRRGDSRGLSPRISYLKSGSGGGLLCPAGHVGIFRAAV 173
Db 1129 CGSSDLVLTVRHADVIPVRRGDSRGLSPRISYLKSGSGGGLLCPAGHVGIFRAAV 1188

Qy 174 STRGVAKAVDFIPVESLETTMRSP 197
Db 1189 CTRGVAKAVDFIPVENLGTMRSP 1212

RESULT 10
Q91RR8 PRELIMINARY; PRT; 181 AA.
AC Q91RR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;

```

```

RN  [1]
RC  SEQUENCE FROM N.A.
RA  STRAIN-Pt.1V;
RT  Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT  "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL  Clinical Strains of the Hepatitis C Virus.";
DR  Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF369235; AAK54560.1; -
DR  InterPro: IPR004109; HCV_NS3.
DR  Pfam: PF02907; HCV_NS3.
DR  NS3 protease.
DR  Hepatitis C virus.
KW  Protease.
FT  NON_TER 1 1
FT  NON_TER 181 181
SQ  SEQUENCE 181 AA; 19130 MW; 85091869299B7C35 CRC64;

Query Match 82.6%; Score 839; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 5.5e-72;
Matches 165; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLTWYVHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCTWYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 11
Q91RT5 PRELIMINARY; PRT; 181 AA.
ID Q91RT5
AC Q91RT5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369218; AAK54543.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR NS3 protease.
DR Hepatitis C virus.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19130 MW; 85D91869299B7C35 CRC64;

Query Match 82.6%; Score 839; DB 12; Length 181;
Best Local Similarity 92.7%; Pred. No. 5.5e-72;
Matches 165; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLTWYVHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCTWYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 12
Q91RR5 PRELIMINARY; PRT; 181 AA.
ID Q91RR5
AC Q91RR5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.3U;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369238; AAK54563.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR NS3 protease.
DR Hepatitis C virus.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match 82.4%; Score 837; DB 12; Length 181;
Best Local Similarity 92.1%; Pred. No. 8.5e-72;
Matches 164; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLTWYVHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCTWYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 13
Q91RR2 PRELIMINARY; PRT; 181 AA.
ID Q91RR2
AC Q91RR2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369241; AAK54566.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR NS3 protease.
DR Hepatitis C virus.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```

```

DB 124 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 12
Q91RR5 PRELIMINARY; PRT; 181 AA.
ID Q91RR5
AC Q91RR5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.3U;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369238; AAK54563.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR NS3 protease.
DR Hepatitis C virus.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19084 MW; 3B5E8161F2100A72 CRC64;

Query Match 82.4%; Score 837; DB 12; Length 181;
Best Local Similarity 92.1%; Pred. No. 8.5e-72;
Matches 164; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 19 TAYAAQTRGEGQGTOKTSHTGRDNQVEGEVQIVSTATOTFLATSIINGVLTWYVHGAGTGT 78
DB 4 TAYAAQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCTWYVHGAGTGT 63
QY 79 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 138
DB 64 IASPKGPVTQMTYNDKDLVGMWAPQGSRLTPTCTGSSDLYLVTRHADVIPVRRGDSR 123
QY 139 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 196
DB 124 GSLLSPRPISYLKSGSGGLPCPAGHAGVIFRAAVSTRGVAKAVDFIPVESLETTMRS 181

RESULT 13
Q91RR2 PRELIMINARY; PRT; 181 AA.
ID Q91RR2
AC Q91RR2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE NS3 protease (Fragment).
OS Hepatitis C virus.
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Pt.4V;
RA Holland-Staley C.A., Kovari L.C., Golenberg E., Mayers D.L.;
RT "Genetic Diversity and response to IFN of the NS3 Protease Gene from
RL Clinical Strains of the Hepatitis C Virus.";
DR Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369241; AAK54566.1; -
DR InterPro: IPR004109; HCV_NS3.
DR Pfam: PF02907; HCV_NS3; 1.
DR NS3 protease.
DR Hepatitis C virus.
KW Protease.
FT NON_TER 1 1
FT NON_TER 181 181
SQ SEQUENCE 181 AA; 19123 MW; 1CAE817345ED809D CRC64;
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 2560.57 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGGVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xl  
-Q=/cqn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151919\_28310/app\_query.fasta\_1.2872  
-DB=GenEmbl -OFMT=faastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.ecgn\_1\_14686/runat\_29082003\_151919\_28310 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pln:\*  
35: em.htg\_rdd:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	882.5	86.9	12734	6	AR179057 Sequence
2	875.5	86.2	1998	6	AR145264 Sequence
3	872.5	85.9	1998	6	AR145268 Sequence
4	871.5	85.8	1998	6	AR145263 Sequence
5	868.5	85.5	651	6	AR145254 Sequence
6	868.5	85.5	1998	6	AR145267 Sequence
7	867.5	85.4	1998	6	AR145262 Sequence
8	865.5	85.2	651	6	AR145258 Sequence
9	864.5	85.1	651	6	AR145253 Sequence
10	864.5	85.1	1998	6	AR145266 Sequence
11	863.5	85.0	1998	6	AR145261 Sequence
12	862.5	85.0	2016	6	AR145269 Sequence
13	862.5	84.9	12734	14	AF268278 Pestivirus
14	861.5	84.8	651	6	AR145257 Sequence
15	860.5	84.7	651	6	AR145252 Sequence
16	860.5	84.7	1998	6	AR145265 Sequence
17	860.5	84.7	2016	6	AR145270 Sequence
18	860	84.6	648	6	AR145274 Sequence
19	858	84.4	648	6	AR145272 Sequence
20	857.5	84.4	651	6	AR145256 Sequence
21	857.5	84.4	651	6	AR145260 Sequence
22	856.5	84.3	651	6	AR145251 Sequence
23	856	84.3	648	6	AR145273 Sequence
24	854	84.1	648	6	AR145271 Sequence
25	853.5	84.0	651	6	AR145255 Sequence
26	853.5	84.0	651	6	AR145259 Sequence
27	851	83.8	8157	6	AR127810 Sequence
28	851	83.8	8157	6	BD081911 Hepatitis
29	849	83.6	1932	6	AR127809 Sequence
30	849	83.6	1932	6	BD081910 Hepatitis
31	848.5	83.5	9646	6	AR110828 Sequence
32	848.5	83.5	9646	6	BD069982 Functiona
33	848.5	83.5	9646	14	AF009606 Hepatitis
34	848.5	83.5	12980	6	AR110831 Sequence
35	848.5	83.5	12980	6	BD069985 Functiona
36	844.5	83.1	5360	6	AR118686 Sequence
37	844.5	83.1	5360	6	I06434 Sequence 48
38	844.5	83.1	5360	6	I09328 Sequence 8
39	844.5	83.1	6785	6	AR118692 Sequence
40	844.5	83.1	6785	6	I06440 Sequence 54
41	844.5	83.1	6785	6	I09329 Sequence 10
42	844.5	83.1	7310	6	AR118696 Sequence
43	844.5	83.1	7310	6	I09331 Sequence 15
44	844.5	83.1	7310	14	HPCPOLYP
45	844.5	83.1	8316	6	AR118703 Sequence

ALIGNMENTS

AR179057  
LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6326137.  
ACCESSION AR179057  
VERSION AR179057.1 GI:20220612  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 12734)  
AUTHORS Hong, Z., Lai, V. C. H. and Lau, J. Y. N.  
TITLE Hepatitis C virus protease-dependent chimeric pestivirus  
JOURNAL Patent: US 6326137-A 1 04-DEC-2001;  
FEATURES Location/Qualifiers  
source 1..12734  
BASE COUNT 4032 a 2604 c 3295 g 2803 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3 77e-64 Length: 12734  
Score: 882.50 Matches: 176  
Percent Similarity: 92.31% Conservative: 4  
Best Local Similarity: 90.26% Mismatches: 12  
Query Match: 86.86% Indels: 3  
DB: 6 Gaps: 1  
US-09-965-594-22 (1-197) x AR179057 (1-12734)  
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 413 GGTAGTGTGTATTTGTTAGTAAATTTTATCTGTTAGTGTAGTATCATCAGCGGTAC 472  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyValArgAspLys 41  
Db 473 GCGGACAGACGAGAGGCGCTCTAGGGTGAAGATCACCAGTCGACTGCGCGGACAAA 532  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 533 AACCAAGTGGAGGGTGAGTCCAGATCGTCACTGCTCAACCAACCTCTCTGGCAACG 592  
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 593 TGCATCAATGGGGTATGTGGACTCTTACCACGGGGCGGAGACGAGACCATCGCATCA 652  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
Db 653 CCCAAGGTCTCTATCCAGATGATACCAATGTGGACCAAGACCTTGTGGCTGGCC 712  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 713 GCTCTCAAGTTCCTCGCTCATGTACACCTCGACCTCGGCTCTCTCGACCTTTACCTG 772  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 773 GTTACGAGCAGCCGAGCTCATTCCTGCGCGCGGAGGTGATAGAGGGTAGCCTG 832  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 833 CTTTCGCCCGCCCATTTCTACCTACCTAAAGGCTCTCTCGGGGGTCCGCTGTGCCCC 892  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
Db 893 GCGGGACACGCGTGGGCGCTATTTCAGGCGCGCGGTGTCACCCGTGGAGTGCCCAAGCG 952  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
Db 953 GTGGACTTTATCCCTGTGGAGAACCTAGACACCAACCATGAGATCC 997  
RESULT 2  
AR145264  
LOCUS AR145264 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 105 from patent US 6211338.

AR145264  
VERSION AR145264.1 GI:15107131  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1998)  
AUTHORS Malcolm, B. A., Taremi, S. Shane., Weber, P. C. and Yao, N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 105 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..1998  
BASE COUNT 411 a 595 c 569 g 423 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.73e-64 Length: 1998  
Score: 875.50 Matches: 167  
Percent Similarity: 92.86% Conservative: 15  
Best Local Similarity: 85.20% Mismatches: 11  
Query Match: 86.17% Indels: 3  
DB: 6 Gaps: 1  
US-09-965-594-22 (1-197) x AR145264 (1-1998)  
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
Db 64 GGTCTGTGTATTGTTAGTAAATTTTATCTGTTAGTGTAGTATCATCAGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyValArgAspLys 41  
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTGCAAGAAGACTAGCTTTACAGCGCGGACAAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
Db 184 AACCAAGTGGAGGGTGGTTCAGGTGGTTTCCACCGCAACAATCTCTCTCGCGACC 243  
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
Db 244 TCGCTCAACGGCGTGTGGTGGACCGTTTACCATTGCTGGCTCAAGACCTTACCGGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
Db 304 CCAAGGGCGCAATACCCAGATGATACATAATGTGGACAGGACCTCTCGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
Db 364 GCGCGCGCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
Db 424 GTCACGAGACATGCTACGCTCATTCGGTGGCGCGCGGCGGAGACAGTAGGGGAGCGCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
Db 484 CTCTCCCCAGGCGCTCTCTCTACTTGAAGGGCTCTCTGGGTGGTCCACTGCTCGCCCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
Db 544 TCGGGGACGCTGTGGCATCTTCCGGGCTGCGGTATGCACCCGGGGGTTTCGGAGGCG 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGCTCTCCG 651  
RESULT 3  
AR145268  
LOCUS AR145268 1998 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 109 from patent US 6211338.  
ACCESSION AR145268  
VERSION AR145268.1 GI:15107135

## KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin B.A., Taremi, S.Shane., Weber, P.C. and Yao, N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 109 03-APR-2001;

FEATURES Location/Qualifiers

source

1..1998

/organism="unknown"

BASE COUNT 411 a 595 c 569 g 423 t

ORIGIN

Alignment Scores:  
 Pred. No.: 3 09e-64 Length: 1998  
 Score: 872.50 Matches: 166  
 Percent Similarity: 92.86% Conservative: 16  
 Best Local Similarity: 84.69% Mismatches: 11  
 Query Match: 85.88% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARL45268 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTATCATCGGCTAC 123  
 |||||  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 |||||  
 Db 124 TCCCAACAGACGGGGGCTACTTGGTGCAGAAAGACTAGCTTACAGCGGGGACAAAG 183  
 |||||  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 AACAGGTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACATCTCTCTGCGGACC 243  
 |||||  
 QY 62 SerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 |||||  
 Db 244 TCGCTCAACGGGGTGTGTTGGACCGTTTACCATGTGCTGCTCAAGACCTTAGCCGGC 303  
 |||||  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 |||||  
 Db 304 CCNAGGGGCCAATACCCAGATGTACACTATGTGGACAGACCTCTGCGCTGGCAG 363  
 |||||  
 QY 102 AlaProGlnCysArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 364 GCGCCCGCGGGCGGCTTCTTGTACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 |||||  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 424 GTCACGAGACATGCTGACGTCTATCCGGTGGCGGGGGGCGACAGTAGGGGAGCCTG 483  
 |||||  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 |||||  
 Db 484 CTCCTCCCGAGGCTCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543  
 |||||  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 |||||  
 Db 544 TCGGGGACAGCTGTGGGACATCTTCCGGGCTGCCGTATGCACCGGGGGGTTCCGAAGGCG 603  
 |||||  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 |||||  
 Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACACTACTATCGGCTCTCCG 651

RESULT 4

ARL45263

LOCUS

DEFINITION Sequence 104 from patent US 6211338.

ACCESSION ARL45263

VERSION ARL45263.1 GI:15107130

KEYWORDS

SOURCE Unknown.

## ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1998)

AUTHORS Malcolin B.A., Taremi, S.Shane., Weber, P.C. and Yao, N.

TITLE Single-chain recombinant complexes of hepatitis C virus NS3

JOURNAL protease and NS4A cofactor peptide

PATENT: US 6211338-A 104 03-APR-2001;

FEATURES Location/Qualifiers

source

1..1998

/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t

ORIGIN

Alignment Scores:  
 Pred. No.: 3 75e-64 Length: 1998  
 Score: 871.50 Matches: 167  
 Percent Similarity: 92.35% Conservative: 14  
 Best Local Similarity: 85.20% Mismatches: 12  
 Query Match: 85.78% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARL45263 (1-1998)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTCTGTTGTTATGTTGGTAGAATTATTTATCTGTTAGTATCATCGGCTAC 123  
 |||||  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 |||||  
 Db 124 TCCCAACAGACGGGGGCTACTTGGTGCATCAAGACTAGCTTACAGCGGGGACAAAG 183  
 |||||  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 AACAGGTCGAGGAGAGGTTCCAGTGGTTTCCACCGCAACACATCTCTCTGCGGACC 243  
 |||||  
 QY 62 SerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 |||||  
 Db 244 TCGCTCAACGGGGTGTGTTGGACCGTTTACCATGTGCTGCTCAAGACCTTAGCCGGC 303  
 |||||  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 |||||  
 Db 304 CCNAGGGGCCAATACCCAGATGTACACTATGTGGACAGACCTCTGCGCTGGCAG 363  
 |||||  
 QY 102 AlaProGlnCysArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 364 GCGCCCGCGGGCGGCTTCTTGTACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
 |||||  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 424 GTCACGAGACATGCTGACGTCTATCCGGTGGCGGGGGGCGACAGTAGGGGAGCCTG 483  
 |||||  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 |||||  
 Db 484 CTCCTCCCGAGGCTCTCTCTACTTGAAGGCTCTTCCGGTGGTCCACTGCTCTGCCCT 543  
 |||||  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 |||||  
 Db 544 TCGGGGACAGCTGTGGGACATCTTCCGGGCTGCCGTATGCACCGGGGGGTTCCGAAGGCG 603  
 |||||  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 |||||  
 Db 604 GTGGACTTTGTGCCCGTAGAGTCCATGGAACACTACTATGCGGCTCTCCG 651

RESULT 5

ARL45254

LOCUS

DEFINITION Sequence 95 from patent US 6211338.

ACCESSION ARL45254

VERSION ARL45254.1 GI:15107121

KEYWORDS

SOURCE Unknown.

ORGANISM

Unclassified.

ARL45254

Sequence 95 from patent US 6211338.

ACCESSION ARL45254

VERSION ARL45254.1 GI:15107121

KEYWORDS

SOURCE Unknown.

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 651)  
 Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 AUTHORS  
 TITLE  
 Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide  
 Patent: US 6211338-A 95 03-APR-2001;

JOURNAL  
 FEATURES  
 source  
 Location/Qualifiers  
 1. .651  
 /organism="unknown"

BASE COUNT 120 a 187 c 200 g 144 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.84e-64 Length: 651  
 Score: 868.50 Matches: 166  
 Percent Similarity: 92.82% Conservative: 15  
 Best Local Similarity: 85.13% Mismatches: 11  
 Query Match: 85.48% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45254 (1-651)

QY 5 GlySerValValIleValGlyAlaGlyIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTTCGTGTTATGTTGGTAGAATATTATCTGCTAGTAGTATCAGCGCTAC 123  
 QY 22 AlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 |||||  
 Db 124 TCCCAACAGACGGGGCGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 ACCAGTCCAGGAGAGAGTTCCAGTGGTTCCACCGCAACACATCTCTCGCGGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 |||||  
 Db 244 TCGTCAACGGCGTGTTCGACCGTTTACATGGTGTGCTGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 |||||  
 Db 304 CCAAGGGGCGCATCACCGAGATGTACATAATGTGGACAGGACCTCGCTGGCGAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 364 CGCGCCCGCGGGCGCTCTCTTGACACCATGTCACCTCTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 424 GTCACGAGACATGTCACGTCATTCCGGTGGCGGGCGGCGGACAGTAGGGGAGCCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 |||||  
 Db 484 CTCCTCCCGGCGCTGTCTCTACTTTGAAGGCTCTGCTGGTGGTCCACTGCTGCGCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 |||||  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTCATGCACCGGGGGGTTCGGAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 |||||  
 Db 604 GTGGACTTTGGCCGTAGAGTCCATGGAATACTACTATGCGGTCT 648

RESULT 6  
 ARI45267  
 LOCUS  
 DEFINITION  
 ARI45267  
 ACCESSION  
 VERSION  
 ARI45267.1 GI:15107134  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unclassified.  
 1 (bases 1 to 1998)  
 Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 AUTHORS

TITLE  
 JOURNAL  
 FEATURES  
 source  
 Location/Qualifiers  
 1. .1998  
 /organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.72e-64 Length: 1998  
 Score: 868.50 Matches: 166  
 Percent Similarity: 92.35% Conservative: 15  
 Best Local Similarity: 84.69% Mismatches: 12  
 Query Match: 85.48% Indels: 3  
 DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45267 (1-1998)

QY 5 GlySerValValIleValGlyAlaGlyIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 |||||  
 Db 64 GGTTCGTGTTATGTTGGTAGAATATTATCTGCTAGTAGTATCAGCGCTAC 123  
 QY 22 AlaGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 |||||  
 Db 124 TCCCAACAGACGGGGCGCTACTTGGTTGCAAGAGACTAGCCTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 |||||  
 Db 184 ACCAGTCCAGGAGAGAGTTCCAGTGGTTCCACCGCAACACATCTCTCGCGGACC 243  
 QY 62 SerIleAsnGlyValLeuThrValThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 |||||  
 Db 244 TCGTCAACGGCGTGTTCGACCGTTTACATGGTGTGCTGCTCAAGACCTTAGCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
 |||||  
 Db 304 CCAAGGGGCGCATCACCGAGATGTACATAATGTGGACAGGACCTCGCTGGCGAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 |||||  
 Db 364 CGCGCCCGCGGGCGCTCTCTTGACACCATGTCACCTCTGGCAGCTCAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
 |||||  
 Db 424 GTCACGAGACATGTCACGTCATTCCGGTGGCGGGCGGCGGACAGTAGGGGAGCCTG 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 |||||  
 Db 484 CTCCTCCCGGCGCTGTCTCTACTTTGAAGGCTCTGCTGGTGGTCCACTGCTGCGCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 |||||  
 Db 544 TCGGGGACGCTGTGGGCATCTTCGGGCTGCGGTCATGCACCGGGGGGTTCGGAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 |||||  
 Db 604 GTGGACTTTGGCCGTAGAGTCCATGGAATACTACTATGCGGTCTCCG 651

RESULT 7  
 ARI45262  
 LOCUS  
 DEFINITION  
 ARI45262  
 ACCESSION  
 VERSION  
 ARI45262.1 GI:15107129  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unclassified.  
 1 (bases 1 to 1998)  
 Malcolm.B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
 AUTHORS  
 TITLE  
 Single-chain recombinant complexes of hepatitis C virus NS3  
 protease and NS4A cofactor peptide

JOURNAL Patent: US 6211338-A 103 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..1998  
/organism="unknown"

BASE COUNT 410 a 596 c 568 g 424 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.16e-64 Length: 1998  
Score: 867.50 Matches: 166  
Percent Similarity: 92.35% Conservative: 15  
Best Local Similarity: 84.63% Mismatches: 12  
Query Match: 85.38% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45262 (1-1998)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTTAGTGTAGTATCAGCGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
|||||
DB 124 TCCCAACACACGGGGGCTACTTGGTTGCAAGATCAGCTTACAGCGCGGACAG 183
|||||
QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 AACCAAGTTCGAGGAGAGGTTCCAGGTGGTTCCACCGCAACAATCCTTCTCGCGACC 243
|||||
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TCGCTCAACGGCGTGTGGACCGTTTACCATGTTGCTGGCTCAAGACCTTAGCCGGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
DB 304 CCAAGGGGCCAATCACCAGATGTACACTATGTGGACAGACCTCGTCGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 364 GCGCCCCCGGGCGGCTTCCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 424 GTCACGAGACATGCTGACGTCATCCGGTGGCGCGGGGGGACAGTAGGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
DB 484 CTCCTCCCGCAGGCTGCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
DB 544 TCGGGCACGCTGTGGGCATCTCCGGGCTGCGGTATGCACCGGGGGGTTCGAAGGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
|||||
DB 604 GTGGACTTGTGCCGTAGAGTCCATGGAACTACTATCGGCTCTCCG 651
```

RESULT 8  
ARI45258  
LOCUS  
DEFINITION Sequence 99 from patent US 6211338.  
ACCESSION ARI45258  
VERSION ARI45258.1 GI:15107125  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 651)  
AUTHORS Malcolim.B.A., Taremi,S.Shane., Weber,P.C. and Yao.N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 99 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..1998  
/organism="unknown"

source 1..651  
/organism="unknown"  
BASE COUNT 120 a 187 c 200 g 144 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3.29e-64 Length: 651  
Score: 865.50 Matches: 165  
Percent Similarity: 92.82% Conservative: 16  
Best Local Similarity: 84.62% Mismatches: 11  
Query Match: 85.19% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x ARI45258 (1-651)

```
QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
|||||
DB 64 GGTTCGTGTTATTTGTTAGTAATTTATTTCTGTTAGTGTAGTATCAGCGCTAC 123
|||||
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
|||||
DB 124 TCCCAACACACGGGGGCTACTTGGTTGCAAGAAAGACTAGCCTTACAGCGCGGACAG 183
|||||
QY 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
|||||
DB 184 AACCAAGTTCGAGGAGAGGTTCCAGGTGGTTCCACCGCAACAATCCTTCTCGCGACC 243
|||||
QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
|||||
DB 244 TCGCTCAACGGCGTGTGGACCGTTTACCATGTTGCTGGCTCAAGACCTTAGCCGGC 303
|||||
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
|||||
DB 304 CCAAGGGGCCAATCACCAGATGTACACTATGTGGACAGACCTCGTCGGCTGGCAG 363
|||||
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
|||||
DB 364 GCGCCCCCGGGCGGCTTCCCTTGACACCATGCACCTGTGGCAGCTCAGACCTTACTTG 423
|||||
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141
|||||
DB 424 GTCACGAGACATGCTGACGTCATTCGGTGGCGCGGGGGGACAGTAGGGGGAGCCTG 483
|||||
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
|||||
DB 484 CTCCTCCCGCAGGCTGCTCTCTACTTGAAGGCTCTGCTGGTGGTCCACTGCTCTGCCCT 543
|||||
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
|||||
DB 544 TCGGGCACGCTGTGGGCATCTCCGGGCTGCGGTATGCACCGGGGGGTTCGAAGGCG 603
|||||
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
|||||
DB 604 GTGGACTTGTGCCGTAGAGTCCATGGAACTACTATCGGCTCT 648
```

RESULT 9  
ARI45253  
LOCUS  
DEFINITION Sequence 94 from patent US 6211338.  
ACCESSION ARI45253  
VERSION ARI45253.1 GI:15107120  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 651)  
AUTHORS Malcolim.B.A., Taremi,S.Shane., Weber,P.C. and Yao.N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4A cofactor peptide  
JOURNAL Patent: US 6211338-A 94 03-APR-2001;  
FEATURES Location/Qualifiers  
source 1..651  
/organism="unknown"

**Alignment Scores:**

NTSTYMO

**Alignment Scores:**

Pred. No.: 1.77e-63 Length: 1998  
Score: 863.50 Matches: 166  
Percent Similarity: 91.84% Conservative: 14  
Best Local Similarity: 84.69% Mismatches: 13  
Query Match: 84.99% Indels: 3  
DB: 6 Gaps: 1

US-09-965-594-22 (1-197) x AR145261 (1-1998)

QY 5 GlySerValValIleValGluValGlnIleValSerThrAlaThrPheLeuAlaThr 21  
|||||  
DB 64 GTTCTGTTGTTATTTGTTGTTAGATTTATTTATCTGTTAGTGTAGTATCATCGGCTTAC 123  
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
:|||||  
DB 124 TCCCAACAGACCGCGGCTACTTGGTTGCATCATCTAGTCTTACAGCGCGGACAAG 183  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrPheLeuAlaThr 61  
|||||  
DB 184 ACCAGGTGAGGAGAGGTTTCAGTGGTTTCCACCGCAACACAACTCTTCTGGCGACC 243  
QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
:|||||  
DB 244 TGGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTGCATCAAGACCTTAGCGCGC 303  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101  
:|||||  
DB 304 CCAAGGGGCAATCACCCAGATGTACACTAATGTGGACCAAGGACCTCGTGGCTGGCAG 363  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 364 GCGCCCCCGGGCGGCTTCTTGACACCATGTGCGACCTGTGGCAGCTCAGACCTTTACTTG 423  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 424 GTCAGGACATGCTGACGTCATTCGCGTGGCGGGCGCGACAGTATGGGGGAGCCTG 483  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
:|||||  
DB 484 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTGGGTGGTCCACTGCTGCGCT 543  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
|||||  
DB 544 TCGGGGACGCTGGGGCATCTTCGGGCTCCGATGACCCCGGGGGTTGCGAAGCG 603  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
:|||||  
DB 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 12  
AR145269 2016 bp DNA linear PAT 08-AUG-2001  
LOCUS  
DEFINITION Sequence 110 from patent US 6211338.  
ACCESSION AR145269  
VERSION AR145269.1 GI:15107136  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2016)  
AUTHORS Malcolim,B.A., Taremi,S.Shane., Weber,P.C. and Yao,N.  
TITLE Single-chain recombinant complexes of hepatitis C virus NS3  
protease and NS4 cofactor peptide  
JOURNAL Patent: US 6211338-A 110 03-APR-2001;  
FEATURES Location/Qualifiers  
Source 1. 2016  
/organism="unknown"

BASE COUNT 412 a 603 c 570 g 431 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.79e-63 Length: 2016  
Score: 863.50 Matches: 166

Percent Similarity: 91.84% Conservative: 14  
Best Local Similarity: 84.69% Mismatches: 13  
Query Match: 84.99% Indels: 3  
DB: 6 Gaps: 1  
US-09-965-594-22 (1-197) x AR145269 (1-2016)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
DB 82 GGTTCTGTTGTTATTTGTTGTTAGATTTATTTATCTGTTAGTGTAGTATCATCGGCTTAC 141  
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
:|||||  
DB 142 TCCCAACAGACCGGGGCTACTTGGTTGCATCATCTAGTCTTACAGCGCGGACAAG 201  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrPheLeuAlaThr 61  
|||||  
DB 202 ACCAGGTGAGGAGAGGTTTCAGTGGTTTCCACCGCAACAACTCTTCTGGCGACC 261  
QY 62 SerIleAsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
:|||||  
DB 262 TGGCTCAACGGCGTGTGTGGACGTTTACCATGGTGTGCTGCCTCAAAGACCTTAGCGCGC 321  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPln 101  
:|||||  
DB 322 CCAAGGGGCAATCACCCAGATGTACACTAATGTGGACCAAGGACCTCGTGGCTGGCAG 381  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
DB 382 GCGCCCCCGGGCGGCTTCTTGACACCATGTGCGACCTGTGGCAGCTCAGACCTTTACTTG 441  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeu 141  
|||||  
DB 442 GTCAGGACATGCTGACGTCATTCGCGTGGCGGGCGCGACAGTATGGGGGAGCCTG 501  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
:|||||  
DB 502 CTCCTCCCGCCAGCGCTGCTCTACTTGAAGGGCTCTTGGGTGGTCCACTGCTGCGCT 561  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
|||||  
DB 562 TCGGGGACGCTGTGGCATCTTCGGGCTGCGGCTATGACCCCGGGGGTTGCGAAGCG 621  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
:|||||  
DB 622 GTGGACTTTGTGCGCGTAGAGTCCATGGAACACTACTATGCGGTCTCCG 669

RESULT 13  
AF268278 12734 bp RNA linear VRL 12-JUL-2000  
LOCUS  
DEFINITION Pestivirus type 1, complete genome.  
ACCESSION AF268278  
VERSION AF268278.1 GI:9049956  
KEYWORDS  
SOURCE Pestivirus type 1  
ORGANISM Pestivirus type 1  
Pestivirus type 1  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Pestivirus.

REFERENCE 1 (bases 1 to 12734)  
AUTHORS Lai,V.C., Zhong,W., Skelton,A., Ingravallo,P., Vassiliev,V.,  
Donis,R.O., Hong,Z. and Lau,J.Y.  
TITLE Generation and characterization of a hepatitis C virus NS3  
protease-dependent bovine viral diarrhea virus  
JOURNAL J. Virol. 74 (14), 6339-6347 (2000)  
MEDLINE 20323484  
PUBMED 10864644

REFERENCE 2 (bases 1 to 12734)  
AUTHORS Lai,V.C.H. and Hong,Z.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research  
Institute, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA  
FEATURES Location/Qualifiers  
Source 1. 12734



/organism="Pestivirus type 1"

/mol\_type="genomic RNA"

/db\_xref="taxon:11099"

1..385

386..12508

/codon\_start=1

/product="polyprotein"

/protein\_id="AF82566.1"

/db\_xref="gi:9049957"

/translation="MELNTNGSGSVVIVGRIVLSGGSSITACAOQTGRLGCKLITSL  
TGRDKNVEGOIVSTATQATLACINGVCHTVHAGTRIASPKGVIOYINVD  
QDLVNPAPQSGSLIPCTCGSSDLVTRHANVIVRRRSGSDSLSPRLSTLKG  
SSQDLPCPAGHAGVLPAAVCTGRVAKAVDFIPVENLETTIRSGGADTDDVCCSM  
SYSDTEEGATKQKDRLEGRKMPKIPKESKDSKTPDPATVIVGVYQVRKK  
GKTSKNTDGLYHNKKNPOESRKLKALAMAIIVLFOVTGENTIQNLQDNG  
TEGTORAMFQVGNVSLHGLPEKICTGVPSHLATDIELKTHGMADSEKNTYCCR  
LORHENKHCWCNWNIEPMLVNMRTQANLTGQPPRECAVTCRVDRADLNVYTOA  
RDSPTLTGCKCKNPSFAGILMRGCPNFEIAASDVLPKHEHRISHPQDTLILVDGL  
TNSLEGARQATKLTWLGQGLGILGKLENKSKTWFGAYASPYCDVDKIGIYWT  
KNCPTACLPKATKIVGPKFDTNAEDGKILHEMGHLSVLLGLSLVLSDFAPETASV  
MYLILFESIQSHVDVMDCKDTQNLNTVELTADVIPGSMNKGKWCIRPNMWPYET  
TVVLAFEVSOVVKVLRALRDLTRINAAATTAFLICLVKIVRGOMVOGILWLLIT  
VGVOHLDCRPEFSYATAKDERIGOLAGELTTTWKEYSPMKLEDTMTWAKEDCKLM  
YLQRCRTRERLAILLHTRALPSTVVPFKLFDGKQEDVVMNDNFEEGLPCDAPIV  
RGFTNGLGPFAPQVCPIGWTVGTSYCSFNMDTLATTIVRTIRSKKPPHRRQGLI  
QKNGEDLHNCILGNNMCTVPGDOLLYKGSIESCKGCGYQFKESGLPHYIGKCL  
ENETGYLDVSCNREGVAVIPQGLKCKIGKTVQVIAMDTKLGPMPCREYETIS  
LGVPEKTACTFNTKTKNRYFEPDSYFOQYMLKEGYOTWFDLEVTDHRRDYFAESI  
LVVVALLGTVYLLVTVMLSEOKALGIOYGSVGMNGLTHNNIEVTVYVFL  
LYLLREESVKKVLLLYLHLYVHPKSVTIIILMIGDVVKADSGOEVYLGKIDLCFT  
TVYLVIGLLIARSDPTVPLVIMAAARVTELTHQPGVDIAVAVTIIILAVSYVTR  
YFRKMWQOCLIRSVLIRSLIYIGRIEMPEVTIPNWRPLTLLILYLSITTVTD  
WKYDVAGLLQDVPILLVTVLWADFLILIPYELVLYLKYLRVTDIERSWLG  
IDTVYDVYDVSDEGVYFPFRQAQGNFSLIPLIKATLISCVSKWOLYMSY  
LITDFMYHRKVIEISGVTYISRLVAALIELNMSBEESKGLKYLKLSGRUN  
LIITKHVYMRVTSYGEVYGMKPIWTIIKATLSKSRHCLICTVCEGRWKWGTG  
PKCRHCPKTCMSGLDPERHYKLFIEGHNFCMCRSCCGKHFEPWDEPKSAR  
YCAECNLHPAEPDQWAESSMLGLKITPALMDGVYDITEWAGCQRVGISPDTHRL  
PCHISFGSRMPOEYNGVQYTAGOLFRLNPLVATIKVMMLNMGNEEGNIEHL  
GWLIRGPAVCKITHEKCHINDLKTAFGIMPRGTTPRAPRPTSLLYKVRGLE  
TGWAYTHQGISYDHYTAGDLVDCSMGRVTVQCSNNRLTDEYGVKTVDSGCPD  
GARCIVNPAVNISSKGAVALHOKIGBFTCTVAGTAPAFDLNKLKAGWSGLPFE  
ASSCRVVRVYKVGKNSKPKTMSGIQTVSKNTADLTENVKITSNRCDDPKOITLA  
TGAGKTTELKAVIEIGRHKRVLLIPIAAAESVYVYARLKHPSISNLAIGDKKE  
GDMAITAYISGFTGMPQKRAAMVEYSIFLDEIHCAPEQLAIKGIHRFSES  
IRVAMTATPAGSVTQKHPITEEFAPEVHKGEDLGSQFLDINGLKIPIVDMKGM  
LVFVPTRMAYEAKKLKARYNSGYISGEDPANRVYTSQSPYIVATNAIESGV  
LPDLDTVIDTCLACERKRVSSKIPFIVTGLKRMVTVGEQARRGRVGRVPGRYR  
SQTATSGDYHYDLQAOARYGEDINTKSPRENNYDWSLYEEDSLITOLEILNN  
LLIISEDPAVKNIIMATDHPETIOLAYNSYEQVPLPKIRNGEVTDYENYFLN  
ARKLGEDVPYIVATEDDLAVLLGLDWDPDGNOQVETGKALKQVTLGLSSNALL  
VALFGYVQALSKRHVPLTDIYTEDQREDITHLOAPNAIKTDGTFETELKELAS  
GDVBEKMGASDAANGGLFEVQAEKIKTAPLFENAEAAKGYOKFIDSLIENKEE  
IIRYGLMGTALYKISIAARLHETAFATLVKLWAPGGESVDHYKQAAVLDVLYVYV  
MKNPSFGDSFETOQGRFVASLFIASALATYTYKTNWNNLSKVWEPALAYLPYATV  
LKMPFTRAAVESLSTIKYKLSIRKSGDLGLTGISAAMEILSONPVSIGVSM  
LGWGAIAHNAIESSEKQTLMLKVFVKNFLDQAAATDELVKENPKIIMALEFAVTI  
GNPLRLYHLXYGYKWEAKELSERTAGRNLTILMEFAPELLGMDSGKIRNLSGN  
YILDIYGLDHQKIRKMGVWLPWAPFSCDMPDSERIRLPTDNLTRIVKLEKGOVE  
EMAFKNVGGKLTNVEGSPFLCNRNPGFVNVYRTYDDNLRILPKVAKLPGCV  
HYKGVATKIDYSGKMLLATDKWEVHGVTIRAKRYTGVGPNFNGAYLGDPEPHRALV  
ERDCATITKNTVQFLKMKGCATYDLTISNIRLIELVHRNLEBEKEIPTATVTWL  
AYTFVNEQDCTIKPVLGERVLPDPVVDINLOPEVOVDTSFVGITITGRETMTGVT  
VLEKVEPDSADNSQVIGLDEGNYPGPGIQTHTLEEHNRDARPFIMLSGRNIS  
NRAKTARNILYTDNREIDLMAAGMLVVALROVDPELSEWDFKTFDLEALE  
ALSJGQPKQVQKVEARYIEQKQVEIPNFASDDPVFELVANKDKRYLIVGDVE  
VKDAQALGATDQTRIIKEVSGTAYMKLSWFLQASNKQMSLTPLFEILLRCPPAT  
KSNKGHMASYQLAAGNWEPLGCGVHLGTIPARRVKIHPYEAVLKLDKTFEBEKKPR  
VKDTPVIREHNKWLKLTIFOGNLTNKKMLNPKLSQOLDREGRKNIYHQIGTINSS  
AGIRLEKLPVRAQTDKTKTHEAIRDKIKSNQNPFLNKKLEIFHTIAQPTLKH  
YGEVTEQLEAGNIRKGAAGFLEKKNIAGEVDEKHLVEQLVRLKAGQPIYETFAI  
PKNEKROVDQMGADLVREKRPRIQYPEAKTFLATIKYNNVNWQOQVPIYGECK  
TPLNFIDKVRKENDSFNEPVAVSFDTKAMDQVTSKDLQLICEIGIYKYYKEMHFKFI

DTITDHMEVPIVITADGEYVIRNQSGSGQDTSAGNSMLNVLNMTYAFCESTGVYK  
SFNEVARIHVCQDDGFLITBKGLGKLPANKQMOILHEAGKPKQITTECKMKVAFREFD  
IEFCSHTPVPVRWSDNTSSHMAGRDATVILSKMATRLDSSGERGTAYEKAVAFSPLL  
MYSNPLVRRLICLLVLSQOPEPDSKHATYYKGDPIGAYKDVIGRNLSELKRTGEK  
LANLKSLSLIGWIKHTSKRIQDCVAIGKEGNMLYNADRLISSKTHGLYIPDGF  
TLQKHVEYEQLOLRTETNPVMGVGTERYKLGPIVNLRLRLKILLMTAVGVSS  
12509..12734

BASE COUNT 4030 a 2608 c 3293 g 2802 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1,83e-62 Length: 12734  
Score: 862.50 Matches: 173  
Percent Similarity: 91.28% Conservative: 5  
Best Local Similarity: 88.72% Mismatches: 14  
Query Match: 84.89% Indels: 3  
DB: 14 Gaps: 1

US-09-965-594-22 (1-197) x AF268278 (1-12734)

QY 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 413 GGTAGTGTGTATTGTTGGTAGAATGTTTATCTGCTAGTGTATACACGGCGTC 472  
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
|||||  
Db 473 GCCCAGCAGCAGCAGAGGCGCTCTAGGTGTAAAGATCACCACTCTGACTGGCGCGGACAAA 532  
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
|||||  
Db 533 AACCAAGTGGAGGGGAGGTCCAGATCGTCACTGTACCCAAACCTTCTCGGCAACG 592  
QY 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
|||||  
Db 593 TGCATCAATGGGTATGCTGGACTGCTTACCACGGGCGCGACAGGACCATCGCATCA 652  
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101  
|||||  
Db 653 CCCAAGGTCTCTCATCATCATGATGATACCAATGTGGACCAAGACCTTGTGGGCTGGCCC 712  
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 713 GCTCTCAAGGTTCCCGCTCATTCACACCCCTGACCTCGCGCTCTCGGACCTTTACTGT 772  
QY 122 ValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerIleu 141  
|||||  
Db 773 GTTAGAGGACAGCCCAAGCGTATTCCTCCGCGCGCGGAGGTAGTACGAGGGGTAGCTG 832  
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 833 CTTTCGCGCGCGCGCTTTCCTTACCTAAAGGCTCTCTGGGGGTCTGCTGTGTGGCCCC 892  
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
|||||  
Db 893 CGGGGACAGCGCGTGGGCTATTACGGCGCGGTGTACACCGCTGGAGTGGCCAAAGCG 952  
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 953 GTGGACTTTATCCCTGTGGGAACCTAGAGACAACTAGAGACAACTAGAGATCC 997

RESULT 14

AR145257

LOCUS

Sequence 98 from patent US 6211338.

AR145257

ACCESSION

AR145257.1 GI:15107124

VERSION

KEYWORDS

Unknown.

Unknown.

ORGANISM

1 (bases 1 to 651)

AUTHORS

Malcolm,B.A., Tarem,S.Shane., Weber,P.C. and Yao,N.

TITLE

Single-chain recombinant complexes of hepatitis C virus NS3

651 bp DNA linear PAT 08-AUG-2001

protease and NS4A cofactor peptide  
Patent: US 6211338-A 98 03-APR-2001;

JOURNAL  
FEATURES

source  
Location/Qualifiers  
1. .651  
/organism="unknown"  
BASE COUNT 119 a 188 c 199 g 145 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.16e-64 Length: 651  
Score: 861.50 Matches: 165  
Percent Similarity: 92.31% Conservatives: 15  
Best Local Similarity: 84.62% Mismatches: 12  
Query Match: 84.79% Indels: 3  
DB: 6

US-09-965-594-22 (1-197) x AR145257 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTTCTGTTGTTATGTTGGTAGAATATTTTATCTGGTAGTGTAGTATCATCGGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlnGlnLysThrSerHisThrGlyArgAspLys 41  
:::|||||  
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTCATCAGACTAGCCTTACAGCGCGGACAAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
:::|||||  
Db 184 AACAGTCTCAGGAGAGAGTTTCAGTGGTTTCCACCGCAACACAATCCTTCCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
:::|||||  
Db 244 TGGCTCAACGCGGTGTTGGACCGTTTACCATGGTGGCTCAAGACCTTAGCCGGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
|||||  
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACGAGGACCTCGTGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 CGCGCCCCCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGGCGCGGGGGCCGACAGTAGGGGAGCCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCCTCCCGCAGCGCTGCTCCTACTTGAAGGCTCTGCTGTGTCCACTGCTCTGCCCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
:::|||||  
Db 544 TCGGGGACAGCTGTGGCATCTTCGGGCTCCCGTATGCACCGCGGGGTTGCCAAGCGC 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGAAACTACTATGCGGTCT 648

RESULT 15

AR145252

LOCUS

AR145252 651 bp DNA linear PAT 08-AUG-2001

DEFINITION Sequence 93 from patent US 6211338.

ACCESSION AR145252

VERSION AR145252.1 GI:15107119

KEYWORDS

SOURCE Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 651)

Malcolm, B.A., Taremi, S., Shane, Weber, P.C. and Yao, N.

Single-chain recombinant complexes of hepatitis C virus NS3

protease and NS4A cofactor peptide

Patent: US 6211338-A 93 03-APR-2001;

JOURNAL

FEATURES  
source  
Location/Qualifiers  
1. .651  
/organism="unknown"  
BASE COUNT 119 a 188 c 199 g 145 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.69e-64 Length: 651  
Score: 860.50 Matches: 165  
Percent Similarity: 92.31% Conservatives: 15  
Best Local Similarity: 84.62% Mismatches: 12  
Query Match: 84.69% Indels: 3  
DB: 6

US-09-965-594-22 (1-197) x AR145252 (1-651)

Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
|||||  
Db 64 GGTTCTGTTGTTATGTTGGTAGAATATTTTATCTGGTAGTGTAGTATCATCGGCGCTAC 123  
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlnGlnLysThrSerHisThrGlyArgAspLys 41  
:::|||||  
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTTCGAAGATCACTAGCCTTACAGCGCGGACAAG 183  
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
:::|||||  
Db 184 AACAGTCTCAGGAGAGAGTTTCAGTGGTTTCCACCGCAACACAATCCTTCCTGGCGACC 243  
Qy 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
:::|||||  
Db 244 TGGCTCAACGCGGTGTTGGACCGTTTACCATGGTGGCTCAAGACCTTAGCCGGC 303  
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101  
|||||  
Db 304 CCAAGGGGCAATCACCAGATGTACACTAATGTGGACGAGGACCTCGTGGCTGGCAG 363  
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
|||||  
Db 364 CGCGCCCCCGGGCGGCTTCTTGACACCATGCACCTGTGGCAGCTCAGACCTTTACTTG 423  
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
|||||  
Db 424 GTCACGAGACATGCTGACGTCATTCCGGTGGCGCGGGGGCCGACAGTAGGGGAGCCTG 483  
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
|||||  
Db 484 CTCCTCCCGCAGCGCTGCTCCTACTTGAAGGCTCTTCCGGTGTGCTCCTCTGCCCT 543  
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
:::|||||  
Db 544 TCGGGGACAGCTGTGGCATCTTCGGGCTCCCGTATGCACCGCGGGGTTGCCAAGCGC 603  
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
|||||  
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGAAACTACTATGCGGTCT 648

Search completed: August 31, 2003, 00:46:41

Job time : 2569.57 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 ; Search time 182.939 Seconds  
(Without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKKGWVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xip  
-Q=/cgn2\_1/USPTO.spool/US09965594/runat\_29082003\_151918\_28302/app\_query.fasta\_1.2872  
-DB=N.Geneseq.19Jun03 -OPMT=fastap -SUFFIX=ing -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -NORM=ext -THR\_SCORE=PCT -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN\_1\_1\_1412.0runat\_29082003\_151918\_28302 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.19Jun03.:

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2004.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1016	100.0	594	21	AAA73334	Hepatitis C virus
2	1010	99.4	594	21	AAA73333	Hepatitis C virus
3	995	97.9	594	21	AAA73332	Hepatitis C virus
4	980	96.5	594	21	AAA73331	Hepatitis C virus
5	963	94.8	594	21	AAA73330	Hepatitis C virus
6	936	92.1	594	21	AAA73329	Hepatitis C virus
7	929	91.4	588	21	AAA73328	Hepatitis C virus
8	902	88.8	588	21	AAA73327	Hepatitis C virus
9	882.5	86.9	12734	24	ABA95615	Chimeric BVDV/HCV
10	875.5	86.2	1998	20	AAH80355	HCV NS4A-NS3 compl
11	872.5	85.9	1998	20	AAH80354	HCV NS4A-NS3 compl
12	871.5	85.8	1998	20	AAH80353	HCV NS4A-NS3 compl
13	868.5	85.5	1998	20	AAH80352	HCV NS4A-NS3 compl
14	868.5	85.5	1998	20	AAH80351	HCV NS4A-NS3 compl
15	867.5	85.4	1998	20	AAH80350	HCV NS4A-NS3 compl
16	865.5	85.2	651	20	AAH80349	HCV NS4A-NS3 compl
17	864.5	85.1	612	25	ABX15706	Anti-viral synthet
18	864.5	85.1	651	20	AAH80344	HCV NS4A-NS3 compl
19	864.5	85.1	1998	20	AAH80357	HCV NS4A-NS3 compl
20	863.5	85.0	1998	20	AAH80352	HCV NS4A-NS3 compl
21	863.5	85.0	2013	20	AAH80360	HCV NS4A-NS3 compl
22	861.5	84.8	651	20	AAH80348	HCV NS4A-NS3 compl
23	860.5	84.7	651	20	AAH80343	HCV NS4A-NS3 compl
24	860.5	84.7	1998	20	AAH80356	HCV NS4A-NS3 compl
25	860.5	84.7	2016	20	AAH80361	HCV NS4A-NS3 compl
26	860	84.6	648	20	AAH80365	HCV NS4A-NS3 compl
27	858	84.4	648	20	AAH80363	HCV NS4A-NS3 compl
28	857.5	84.4	650	20	AAH80347	HCV NS4A-NS3 compl
29	857.5	84.4	651	20	AAH80351	HCV NS4A-NS3 compl
30	856.5	84.3	651	20	AAH80342	HCV NS4A-NS3 compl
31	854	84.1	648	20	AAH80362	HCV NS4A-NS3 compl
32	853.5	84.0	650	20	AAH80346	HCV NS4A-NS3 compl
33	853.5	84.0	651	20	AAH80350	HCV NS4A-NS3 compl
34	851	83.8	8145	20	AAH23259	Plasmid pRT-BS(+)
35	849	83.6	1933	20	AAH23258	HCV NS3 DNA. Hepa
36	848.5	83.5	9646	19	AAV59361	Hepatitis C virus
37	848.5	83.5	9646	24	ABR87285	cdNA encoding hepa
38	848.5	83.5	12980	19	AAV59364	Hepatitis C virus
39	848.5	83.5	12980	24	ABR87286	Hepatitis C virus
40	848.5	83.5	16622	21	AAZ36212	Nucleotide sequenc
41	844.5	83.1	5300	10	AAH92097	Combined open read
42	844.5	83.1	5360	10	AAH90327	Hepatitis C virus
43	844.5	83.1	6905	10	AAH92103	Combined open read
44	844.5	83.1	7310	10	AAH92106	Combined open read
45	844.5	83.1	7310	10	AAH90336	Composite hepatiti

ALIGNMENTS

RESULT 1	
AAA73334	
ID	AAA73334 standard; DNA; 594 BP.
XX	AAA73334;
AC	
XX	
DT	19-DEC-2000 (first entry)
XX	
DE	Hepatitis C virus NS4A-NS3 fusion protease coding sequence #7.
XX	
KW	Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW	liver failure; liver cancer; mutant; muten; ds.
XX	
OS	Hepatitis C virus.
OS	Synthetic.
XX	
FH	Key Location/Qualifiers



XX SQ Sequence 594 BP; 104 A; 191 C; 152 G; 147 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,44e-85 Length: 594  
 Score: 1010.00 Matches: 196  
 Percent Similarity: 99.49% Conservatives: 0  
 Best Local Similarity: 99.49% Mismatches: 1  
 Query Match: 99.41% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAATAAAAGGATCGGTGTTATCGTCCGCCGATCAACCTGCGGTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyClnGlnGlyThrSerHisThrGlyArgAsp 40  
 DB 61 TACGCTCAGCAGCTCGAGGTGAGCAGGTTGCCAGAGACCTCCACACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAACACAGGTGAAGTGAAGTTCAGATCGTTCCACCGCTACCCAGACCTTCCTGGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTCATCAACGGGTGTTCTGTGGACCGTTTACCACGGTGTGCTACCCGATCCATCGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 DB 241 TCCCGAAGGTCCCGTTACCCAGATGTACCAACGTTGACAAAGACCTGTTGGTTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CAGGCTCCGACGGTTCGGTTCCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
 DB 361 CTGGTTACCCGTCACCGCTGACGTTATCCCGGTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 421 CTGCTGTCCCGCTCCGATCTCTACCTGGAAGGTTCCTCCGGTGGTCCGCTGCTGTC 480  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 481 CCGGCTGTGTCACGCTGTTGGTATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 541 GCTGTTGACTTCATCCCGGTGAATCCCTGGAAACACCATGCTGCTCCCG 591

RESULT 3

AAA73332

ID AAA73332 standard; DNA; 594 BP.

XX AC

XX AA73332;

XX DT

19-DEC-2000 (first entry)

XX DE

Hepatitis C virus NS4A-NS3 fusion protease coding sequence #5.

XX KW

Hepatitis; NS3 protease; viral replication; chronic liver disease;

XX KW

liver failure; liver cancer; mutant; mutein; ds.

XX OS

Hepatitis C virus.

OS Synthetic.

XX FH

Location/Qualifiers

XX CDS

1..594

FT FT

/product= \*NS4A-NS3 fusion protein #5\*

XX FN WO2000040707-A1.  
 XX PD 13-JUL-2000.  
 XX PF 06-JAN-2000; 2000WO-US00345.  
 XX PR 08-JAN-1999; 99US-0115271.  
 XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX WPI; 2000-465976/40.  
 DR P-PSDB; AAB15223.  
 XX PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 .  
 XX Claim 26; Fig 15; 66pp; English.  
 CC The present sequence is the coding sequence for a mutated version of a  
 CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
 CC protease enzymes. These proteins are both essential for the replication  
 CC of the virus, acting to cleave its replicative proteins from the  
 CC polyprotein produced from the HCV genome. Inhibitors of the two proteins  
 CC should be effective as antiviral treatments of HCV infection. This is  
 CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
 CC failure and liver cancer. The present invention concerns a number of NS3  
 CC mutants and NS3-NS4A fusion proteins which can be used to identify  
 CC inhibitors of this type, as well as enabling structural studies of the  
 CC protease and protease-inhibitor complexes. The protein produced from this  
 CC sequence contains the alpha-helix0-1 variant.  
 SQ Sequence 594 BP; 105 A; 189 C; 153 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 3,64e-84 Length: 594  
 Score: 995.00 Matches: 193  
 Percent Similarity: 98.98% Conservatives: 2  
 Best Local Similarity: 97.97% Mismatches: 2  
 Query Match: 97.93% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73332 (1-594)

QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAATAAAAGGATCGGTGTTATCGTCCGCCGATCAACCTGCTCCGTTGACACCGCT 60  
 QY 21 TyrAlaGlnGlnThrArgGlyClnGlnGlyThrSerHisThrGlyArgAsp 40  
 DB 61 TACGCTCAGCAGCTCGAGGTGAGGAGGTGCCAAGAACTCCACAGACCGGTGCTGAC 120  
 QY 41 LysAsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 121 AAAACACAGGTGAAGTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTGCT 180  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 181 ACCTCATCAACGGGTGTTCTGTGGACCGTTTACCACCGTGTGCTGCTGCTGCTGCTGCT 240  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100  
 DB 241 TCCCGAAGGTCCCGTTACCCAGATGTACCAACGTTGACAAAGACCTGTTGGTTGG 300  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 301 CAGGCTCCGACGGTTCGGTTCCCTGACCCCGTGCACCTGCGGTTCCTCCGACCTGTAC 360  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140







```
Percent Similarity: 93.91% Conservative: 0
Best Local Similarity: 93.91% Mismatches: 12
Query Match: 92.13% Indels: 0
DB: 21 Gaps: 0

US-09-965-594-22 (1-197) x AAA73329 (1-594)
QY 1 MetLysLysLysGlySerValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAGGATCCGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTCTGCTGGGTGTCATCATCCTCCCTGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrPheLeuAla 60
DB 121 AAAAACAGGTTGAAGTGAAGTTCAATGTTCCACCGCTGTCAGACCTTCCTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuTyrThrValThrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGTCATCAACGCTGTTGTCGACCGTTTACCACGGTGTGTATCCCGTACCACGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGGAAAGTCCGGTTATCCAGATGATACCAACGTTGACAAAGACCTGGTGGTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CCGGCTCCGACGGTTCCTGCTCCACCGCGTGCACCTGCGGTTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGTTTACCGCTACGCTGAGGTATCCCGGTTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGCCCCCGTCCGATCTCTACTGAAAGGTTCTCCGCGTGGTCCGCTGCTGCTG 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGCTACGCTGTTGGTATCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 191 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCAATCCCGGTTGAAATCCCTGGAACACCACTGCGGTTCCCGG 591

RESULT 7
ID AAA73329
XX AAA73329 standard; DNA; 588 BP.
AC AAA73329;
XX
DT 19-DEC-2000 (first entry)
XX
DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.
XX
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;
KW liver failure; liver cancer; mutant; mutin; ds.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH 1..588
FT CDS
FT /tag= a
FT /product= "NS4A-NS3 fusion protein #2"
XX
PN W0200040707-A1.
XX
PD 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00345.
PF
```

xx  
PR 08-JAN-1999; 99US-0115271.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
XX WPI; 2000-465976/40.  
DR P-PSDB; AAB15220.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
XX  
XX  
PS Claim 26; Fig 12; 66pp; English.  
XX  
XX The present sequence is the coding sequence for a mutated version of a  
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
CC protease enzymes. These proteins are both essential for the replication  
CC of the virus, acting to cleave its replicative proteins from the  
CC polypeptide produced from the HCV genome. Inhibitors of the two proteins  
CC should be effective as antiviral treatments of HCV infection. This is  
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
CC failure and liver cancer. The present invention concerns a number of NS3  
CC mutants and NS3-NS4A fusion proteins which can be used to identify  
CC inhibitors of this type, as well as enabling structural studies of the  
CC protease and protease-inhibitor complexes. The protein produced from this  
CC sequence contains the alpha-helix0-1 variant.  
XX  
SQ Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;

Alignment Scores:  
Pred. No.: 5,38e-78 Length: 588  
Score: 929.00 Matches: 183  
Percent Similarity: 94.42% Conservative: 3  
Best Local Similarity: 92.89% Mismatches: 9  
Query Match: 21 Indels: 2  
DB: 1 Gaps: 1

US-09-965-594-22 (1-197) x AAA73329 (1-588)

QY 1 MetLysLysLysGlySerValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
DB 1 ATGAAAGGATCCGTTGTTATCGTCGGCGGTATAGTACTGACCGT-----GCT 54  
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrSerHisThrGlyArgAsp 40  
DB 55 TACGCTCAGCAGACTCGAGGTGAGGAGGTTCGCCAAGAACCTCCACAGCGGTGCTGAC 114  
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
DB 115 AAAAACAGGTTGAAGTGAAGTTCAATGTTCCACCGCTGTCAGACCTTCCTGGCT 174  
QY 61 ThrSerIleAsnGlyValLeuTyrThrValThrHisGlyAlaGlyThrArgThrIleAla 80  
DB 175 ACCTGTCATCAACGCTGTTGCTGGACCGTTTACCACGGTGTGTTGTTACCGTACCACCTGCT 234  
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
DB 235 TCCCGGAAAGGTCGCGTTTATCCAGATGTACACCAACGTTGACAAAGACCTGCTGGTGG 294  
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
DB 295 CCGGCTCCCGAGGTTCCCGTTCCCGTCCCGGTCGACCTGCGGTTCTCCGACCTGTAC 354  
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140  
DB 355 CTGGTTACCGGTACCGTACCGTTATCCCGGTTCTGCTGCTGGTGACTCCCGTGGTTC 414  
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
DB 415 CTGCTGTCGCCGCTCCGATCTCTACTGAAAGGTTCTCCCGGTGGTCCCGTGGTGGTGC 474



QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 475 CCGGTGGTGCACGCTGTTGGTATCTCCGTGCTGCTGTTGCACCCGGGGTGGCTTAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCCCGG 585

RESULT 8  
 AAA73328  
 ID AAA73328 standard; DNA; 588 BP.  
 AC AAA73328;  
 DT 19-DEC-2000 (first entry)  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..588  
 FT /tag= a  
 FT /product= "NS3-NS4A fusion protein"  
 XX  
 PN W0200040707-AL.  
 XX  
 XX 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 XX  
 XX 08-JAN-1999; 99US-0115271.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 PI  
 XX  
 XX WPI: 2000-465976/40.  
 DR P-PSDB; AAB15212.  
 DR  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 XX Disclosure; Fig 10; 66pp; English.  
 XX  
 CC The present sequence is the coding sequence for a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.  
 XX  
 SQ Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,81e-75 Length: 588  
 Score: 902.00 Matches: 181  
 Percent Similarity: 92.39% Conservative: 1  
 Best Local Similarity: 91.88% Mismatches: 13  
 Query Match: 88.78% Indels: 2  
 DB: 21 Gaps: 1

US-09-965-594-22 (1-197) x AAA73328 (1-588)  
 QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20  
 DB 1 ATGAAAAAAGGTTCCGTTGTTATCGTCGCGCGTATAGTACTGAACGGT-----GCT 54  
 QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnThrGlnLysThrSerHisThrGlyArgAsp 40  
 DB 55 TAGCTCAGCAGACTCGAGGCTGCTGGGTGCTGCTCATCATCCCTCCGTCGCGTGTGAC 114  
 QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60  
 DB 115 AAAAACCCAGGTGGAAGTGAAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTCTGGCT 174  
 QY 61 ThrSerIleAsnGlyValLeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAla 80  
 DB 175 ACCTGCATCAACGGTGTTCCTGACCGGTTTACACCGGTCGTGCTACCCGTCACCATCGCT 234  
 QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100  
 DB 235 TCCCGAAAGGTCGGTTATCCAGATGTACACCAACGTTGACAAAGACCTGGTGTGGTGG 294  
 QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120  
 DB 295 CCGGCTCCGCGAGGTTCCCGTTCCTGACCCCGTGCACCTGCGGTTCTCCGACCTGTAC 354  
 QY 121 LeuValThrArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySer 140  
 DB 355 CTGGTTACCGCTCAGCTGACGTTATCCCGGTCGTCGTCGTGCTGCTGCTGCTGCTGCTG 414  
 QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160  
 DB 415 CTGCTGTCCCGCGTCGCGATCTCCTACTGAAAGGTTCTCCGCTGCTGCTGCTGCTGCTG 474  
 QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180  
 DB 475 CCGGCTGGTCACGCTGTTGGTATCTTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTAA 534  
 QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 535 GCTGTTGACTTCATCCCGGTTGAATCCCTGGAACACCATGCGTTCCCGG 585

RESULT 9  
 ABA95615  
 ID ABA95615 standard; DNA; 12734 BP.  
 XX  
 AC ABA95615;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Chimeric BVDV/HCV NS3-wt sequence.  
 XX  
 KW Pestivirus; Npro; protease; NS3; screening; ds.  
 XX  
 OS Chimeric - Bovine viral diarrhea virus.  
 OS Chimeric - Hepatitis C virus.  
 XX  
 PN USG326137-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 XX 25-JUN-1999; 99US-0344456.  
 XX  
 XX 25-JUN-1999; 99US-0344456.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Hong 2, Lai VCH, Lau JYN;  
 XX  
 DR WPI; 2002-121103/16.  
 XX  
 PT Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)

PT pestivirus genome where the Npro protease gene is replaced with NS3  
 PT protease gene, useful for in vivo screening of compounds which inhibit  
 XX HCV infection

PS Example 2: Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a  
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
 CC each junction site recognised by the Npro protease is replaced with a  
 CC junction site recognised by the HCV NS3 protease. The construct is useful  
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
 CC protease, where screening may be in cell culture or in an animal model.  
 CC The present sequence is a chimeric clone of BVDV (bovine viral diarrhoea  
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.

XX Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;

Alignment Scores:  
 Pred. No.: 5,54e-72 Length: 12734  
 Score: 882.50 Matches: 176  
 Percent Similarity: 92.31% Conservative: 4  
 Best Local Similarity: 90.26% Mismatches: 12  
 Query Match: 86.86% Indels: 3  
 DB: 24 Gaps: 1

US-09-965-594-22 (1-197) x ABA95615 (1-12734)

QY 5 GlySerValIleValGlyArgGlyAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 413 GGTAGTGTGTTATTTGTTGGTAGAATGTTTNTCTGGTAGTGTAGTATCATCGCGGTAC 472  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 473 GCCCAGCAGCAGAGAGGCTCTAGGGTGAAGATCACCAGTCTGACTGGCCGGGACAAA 532  
 QY 42 AsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 533 AACCAAGTGGAGGGTGGAGTCCAGATGTCGTAACCTGCTACCAACCTTCTCGCAACG 592  
 QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 593 TGCATCAATGGGTATGCTGGAGTCTCTACCAAGGCGGACGAGGACCATCGCATCA 652  
 QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrPheGln 101  
 DB 653 CCCAAGGGTCTCTGTCATCCAGATGTATACCAATGTGGACCAAGACCTTGTGGGCTGCC 712  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 713 GCTCTCAAGTTCCTCCGCTCATGTACCCCTGACCTCGGCTCTCTCGSACCTTTACCTG 772  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyArgGlyAspSerArgGlySerLeu 141  
 DB 773 GTTACAGGACCGCGAGCTATCTCCGTCGCCGCGAGGTGATAGCAGGGTAGCCCTG 832  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 833 CTTTCGCCCGCGCCATTTCTTACCTAAAGGCTCTCGGGGGGTCCCGCTGTGTGCC 892  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181  
 DB 893 GCGGGACACCCCTGGCCCTATTACGGCGCGGGGTGTGCACCCGTGGAGTGGCCAAAGCG 952  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196  
 DB 953 GTGGACTTTATCCCTGTGGAGAACCCTAGAGACAACCATGAGATCC 997

RESULT 10

AA80355

ID AAX80355 standard; cDNA; 1998 BP.

XX

AC AAX80355;  
 XX 07-SEP-1999 (first entry)  
 DT  
 DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:105.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.

OS Synthetic.

PN W09928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Disclosure; Page 166-169; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

XX Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;

Alignment Scores:  
 Pred. No.: 2.5e-72 Length: 1998  
 Score: 875.50 Matches: 167  
 Percent Similarity: 92.86% Conservative: 15  
 Best Local Similarity: 85.20% Mismatches: 11  
 Query Match: 86.17% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80355 (1-1998)

QY 5 GlySerValIleValGlyArgGlyAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GGTCTGTGTTATTTGTTGGTAGAATATTATTTATCTGGTAGTGTAGTATCATCGCGGTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACGCGGGGCTTACTTGTTCGAAGAGACTAGCTTACAGCGCGGACAAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTGGAGGGTTCAGTGGTTCACCGCGCAACAAATCCCTTCCTGGCGACC 243

QY 62 SerIleAsnGlyValLeuTyrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TGGCTCAACGGCGTGTGTTGGACCGTTTACCATGGGTGGCTCAAGAACCTTATGCGGCG 303

```
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACAGGACCTGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCCGGCGCGTGTCTGCACATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCCAGAGACATGTGAGCTATTCGGGCTGCCGTATCCACCCGGGGGTTCGGAAGCG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCCAGCCCTGCTCTCTACTTTGAAGGCTCTTCGGGTGGTCCACTGCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACAGCTGTGGGCATCTTCCGGGCTGCCGTATCCACCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCCGTAGAGTCCATGGGAACACTACTATCGCGTCTCCG 651

RESULT 11
ID AAX80359 standard; cDNA: 1998 BP.
XX
AC AAX80359;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:109.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
XX
XX 10-JUN-1999.
XX
XX 24-NOV-1998; 98WO-US24528.
XX
XX 28-JUL-1998; 98US-0094331.
XX
XX 28-NOV-1997; 97US-0067315.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Malcolm BA, Taremi SS, Weber PC, Yao N;
XX
XX WPI; 1999-385385/32.
XX
XX New hepatitis C virus covalent complexes
XX
XX Disclosure; Page 179-182; 21lpp; English.
XX
XX The present invention describes a covalent hepatitis C virus (HCV)
XX NS4A-NS3 complex comprising a central hydrophobic domain of native HCV
XX NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
XX hydrophobic domain of native HCV NS4A peptide is tethered by the linker
XX to the amino terminus of the HCV NS3 protease domain. The present
XX sequence encodes an example of the above complex. The covalent
XX NS4A-NS3 complexes are useful for structural determination and
XX determination of mode of binding of HCV inhibitors by NMR spectroscopy.
XX They can also be used for detecting inhibitors of the protease activity,
XX the helicase activity and the ATPase activity of NS3. The covalent
XX NS4A-NS3 complexes are more soluble, stable and active than the non-
XX covalent protease-peptide complexes previously available.
XX
```

```
SQ Sequence 1998 BP; 411 A; 595 C; 569 G; 423 T; 0 other;
Alignment Scores:
Pred. No.: 4,77e-72 Length: 1998
Score: 872.50 Matches: 166
Percent Similarity: 92.86% Conservative: 16
Best Local Similarity: 84.69% Mismatches: 11
Query Match: 85.88% Indels: 3
DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80359 (1-1998)
QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
DB 64 GGTTCGTGTTTATTTGTTAGTAATATTTTATCTGGTAGTATCATCGGCTTAC 123
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
DB 124 TCCCAACAGACGCGGGGCCCTACTTGGTTGCAAGAACACTAGCCTTACAGCGCGGACAAG 183
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
DB 184 AACGAGTCCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACAAATCCTTCTCGGCGACC 243
QY 62 SerIleAsnGlyValLeuThrPheValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
DB 244 TGGCTCAACGGCGTGTGTGGACCGTTTACCATGTGTGGCTCAAGACCTTAGCGCGC 303
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGln 101
DB 304 CCAAGGGGCCAATCACCAGATGTACACTAATGTGGACAGGACCTGCTGGCTGGCAG 363
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
DB 364 GCGCCCGCCGGCGCGTGTCTGCACATGCACCTGTGGCAGCTCAGACCTTTACTTG 423
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
DB 424 GTCCAGAGACATGTGAGCTATTCGGGCTGCCGTATCCACCCGGGGGTTCGGAAGCG 483
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
DB 484 CTCCTCCCGCCAGCCCTGCTCTCTACTTTGAAGGCTCTTCGGGTGGTCCACTGCTGCCCT 543
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
DB 544 TCGGGGACAGCTGTGGGCATCTTCCGGGCTGCCGTATCCACCCGGGGGTTCGGAAGCG 603
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 604 GTGGACTTGTGCCGTAGAGTCCATGGGAACACTACTATCGCGTCTCCG 651

RESULT 12
AAX80354
ID AAX80354 standard; cDNA: 1998 BP.
XX
AC AAX80354;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:104.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO9928482-A2.
XX
XX 10-JUN-1999.
XX
```

PF 24-NOV-1998; 98WO-US24528.  
 XX 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 XX Malcolml BA, Taremi SS, Weber PC, Yao N;  
 PI WPI; 1999-385385/32.  
 DR New hepatitis C virus covalent complexes  
 XX Disclosure; Page 163-166; 211pp; English.  
 XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.  
 XX Sequence 1998 BP: 410 A; 596 C; 568 G; 424 T; 0 other;  
 SQ

Alignment Scores:  
 Pred. No.: 5,91e-72 Length: 1998  
 Score: 871.50 Matches: 167  
 Percent Similarity: 92.35% Conservative: 14  
 Best Local Similarity: 85.20% Mismatches: 12  
 Query Match: 85.78% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80354 (1-1998)

QY 5 GlySerValIleValGluValArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GTTCTGTTGTTATGTTGGTAGAATATTTTCTGCTAGTGTAGTATCATCGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41  
 DB 124 TCCCAACAGACGCGGGGCTACTTGGTGGCTCAAGACTACGCCCTTACAGCGCGGACAAG 183  
 QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61  
 DB 184 AACCAAGTTCGAGGAGAGGTTTCAGGTGTTTCCACCGCAACACATCTCTCTGCGGACC 243  
 QY 62 SerIleAsnGlyValLeuThrPheValThrHisGlyAlaGlyThrArgThrIleAlaSer 81  
 DB 244 TCGCTCAACGCGGCTGTGTGGACGCTTACCATGCTGCTCAAGACCTTAGCCGCGC 303  
 QY 82 ProLysGlyProValThrGlnMetThrThrAsnValAspLysAspLeuValGlyTyrGln 101  
 DB 304 CCAAGGGGCGCAATCACCAGATGTACACTAATGTGGACCGACCGCTCGTGGCTGGCAG 363  
 QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121  
 DB 364 GGGCCCCCGGGGCGGCTTCTGACACCATGTCACCTGTGCGCTAGACCTTTACTTG 423  
 QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141  
 DB 424 GTCCAGAGACATGCTGACGCTCATTCGCTGCGCGCGGCGGCGACAGTAGGGGCGGCGT 483  
 QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161  
 DB 484 CFTCCCCCGAGCCCTGCTCTCTACTTGAAGGGCTCTTCGGGTGTCCACTCTCTGCCCT 543  
 QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181

DB 544 TCGGGGCACGCTGTGGCATCTTCGGGCTGCGGTATGCACCCGGGGTTCCGAAGCG 603  
 QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197  
 DB 604 GTGGACTTTGTGCCGTAGATCCATGGAACACTACTATGCGGTCTCCG 651

RESULT 13  
 AAX80345  
 ID AAX80345 standard; cDNA; 651 BP.  
 XX AAX80345;  
 XX 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:95.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.  
 XX Hepatitis C virus.  
 OS Synthetic.  
 XX WO9928482-A2.  
 XX 10-JUN-1999.  
 XX 24-NOV-1998; 98WO-US24528.  
 XX 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 XX Malcolml BA, Taremi SS, Weber PC, Yao N;  
 PI WPI; 1999-385385/32.  
 DR New hepatitis C virus covalent complexes  
 PT Disclosure; Page 147-148; 211pp; English.  
 PS The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.  
 XX Sequence 651 BP; 120 A; 187 C; 200 G; 144 T; 0 other;  
 SQ

Alignment Scores:  
 Pred. No.: 2.8e-72 Length: 651  
 Score: 868.50 Matches: 166  
 Percent Similarity: 92.82% Conservative: 15  
 Best Local Similarity: 85.13% Mismatches: 11  
 Query Match: 85.48% Indels: 3  
 DB: 20 Gaps: 1

US-09-965-594-22 (1-197) x AAX80345 (1-651)

QY 5 GlySerValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21  
 DB 64 GTTCTGTTGTTATGTTGGTAGAATATTTTCTGCTAGTGTAGTATCATCGGCCTAC 123  
 QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41

```

Db 124 TCCCAACAGACGCGGGCTACTTGGTTCAGAAAGACTAGCCTTACAGCGCGGACAAAG 183
Qy 42 AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTGAGTGGTTCCACCCGCAACACAAATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGGCGTGTGTGGACCGTTTACCATGTGTGCTGCATCAAGACCTTAGCCGGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATCACCCAGATGTACACTAATGTGGACCAAGACCTCTGGCTGGCGAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GGGCCCCCGGGCGGTCTCTTGACACCATGCCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGACATGCTGACGCTATTCGGTGTGCGCGGCGGCGGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCGCCCGCAGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTGCCACTCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGCACGCTGTGGGCATCTTCCGGCTGCGCTATGCAACCCGGGGGTGGAGGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 648

RESULT 14
AAx80358
ID AAX80358 standard; cDNA; 1998 BP.
XX
AC AAX80358;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:108.
XX
KW HCV; hepatitis C virus; single chain recombinant complex; linker;
KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;
KW hydrophobic domain; covalent complex; detection; inhibitor; ss.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
PN W09928482-A2.
XX
PD 10-JUN-1999.
XX
PF 24-NOV-1998; 98NO-US24528.
XX
PR 28-JUL-1998; 98US-0094331.
XX
PR 28-NOV-1997; 97US-0067315.
XX
PA (SCHE ) SCHERING CORP.
XX
PI Malcolm BA, Taremi SS, Weber PC, Yao N.
XX
PI WPI; 1999-385385/32.
XX
PT New hepatitis C virus covalent complexes
XX
PS Disclosure; Page 176-179; 21lpp; English.
XX
CC The present invention describes a covalent hepatitis C virus (HCV)
CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV

```

```

CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the
CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker
CC to the amino terminus of the HCV NS3 protease domain. The present
CC sequence encodes an example of the above complex. The covalent
CC NS4A-NS3 complexes are useful for structural determination and
CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.
CC They can also be used for detecting inhibitors of the protease activity,
CC the helicase activity and the ATPase activity of NS3. The covalent
CC NS4A-NS3 complexes are more soluble, stable and active than the non-
CC covalent protease-peptide complexes previously available.
XX
SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;
Alignment Scores:
Pred. No.: 1,13e-71 Length: 1998
Score: 868.50 Matches: 166
Percent Similarity: 92.35% Conservative: 15
Best Local Similarity: 84.69% Mismatches: 12
Query Match: 85.48% Indels: 3
DB: 20 Gaps: 1
US-09-965-594-22 (1-197) x AAX80358 (1-1998)
Qy 5 GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 64 GGTTCCTGTTGTTATGTTGTTAGTAATATTATTTATCTGTTAGTGTAGTATCATCGGCTAC 123
Qy 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 124 TCCCAACAGACGCGGGCGCTACTTGGTGTGCATCAAGACTAGCCTTACAGCGCGGACAAAG 193
Qy 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 184 AACCAAGTCGAGGAGAGGTTTCCAGTGGTTTCCACCCGCAACACAAATCTCTCTGGCGACC 243
Qy 62 SerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 244 TCGGTCAACGGCGTGTGTGGACCGTTTACCATGTGTGCTGCATCAAGACCTTAGCCGGC 303
Qy 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrpGln 101
Db 304 CCAAGGGGCCAATCACCCAGATGTACACTAATGTGGACCAAGACCTCTGGCTGGCGAG 363
Qy 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 364 GGGCCCCCGGGCGGTCTCTTGACACCATGCCTGTGGCAGCTCAGACCTTTACTTG 423
Qy 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 424 GTCAGGACATGCTGACGCTATTCGGTGTGCGCGGCGGCGGACAGTAGGGGAGCCTG 483
Qy 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 484 CTCGCCCGCAGCGCTGCTCCTACTTGAAGGGCTCTGCTGGTGTGCCACTCTCTGCCCT 543
Qy 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 544 TCGGGGCACGCTGTGGGCATCTTCCGGCTGCGCTATGCAACCCGGGGGTGGAGGCG 603
Qy 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
Db 604 GTGGACTTTGTGCGCGTAGAGTCCATGGAAACTACTATGCGGTCTCCG 651

RESULT 15
AAx80353
ID AAX80353 standard; cDNA; 1998 BP.
XX
AC AAX80353;
XX
DT 07-SEP-1999 (first entry)
XX
DE HCV NS4A-NS3 complex encoding cDNA SEQ ID NO:103.
XX

```

KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; ss.

OS Hepatitis C virus.  
 OS Synthetic.

XX W09928482-A2.  
 PN W09928482-A2.  
 XX 10-JUN-1999.

PD 24-NOV-1998; 98WO-US24528.  
 PF 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.  
 PA Malcolm RA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.  
 DR New hepatitis C virus covalent complexes

PS Disclosure; Page 160-162; 21pp; English.  
 XX The present invention describes a covalent hepatitis C virus (HCV)

CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The present  
 CC sequence encodes an example of the above complex. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the non-  
 CC covalent protease-peptide complexes previously available.

XX SQ Sequence 1998 BP; 410 A; 596 C; 568 G; 424 T; 0 other;

# Alignment Scores:

Pred. No.:	1,4e-71	Length:	1998
Score:	867.50	Matches:	166
Percent Similarity:	92.35%	Conservative:	15
Best Local Similarity:	84.69%	Mismatches:	12
Query Match:	85.38%	Indels:	3
DB:	20	Gaps:	1

US-09-965-594-22 (1-197) x AAX80353 (1-1998)

QY	5	GlySerValValIleValGlyArgIleAsnLeuSerGlyAsp-----ThrAlaTyr	21
DB	64	GGTTCTGTGTATTGTTGGTAGAATATTTATCTGTTAGTGTAGTATACCGGCTAC	123
QY	22	AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys	41
DB	124	TCCCAACACACCGCGGGCTACTGGTTGCAAGATCCTAGCCTACAGCGCGGACAAG	183
QY	42	AsnGlnValGluGlyValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr	61
DB	184	AACAGGTCGAGGAGAGGCTCAGGTGTTCCACCGCAACAATCTCTCGCGGACC	243
QY	62	SerIleAsnGlyValLeuThrPheValThrValThrGlyAlaGlyThrArgThrIleAlaSer	81
DB	244	TGCGTCAACGGCGTGTGTTGGACCCCTTACCATGCTGCTGCTCAAGACCTTAGCCGCG	303
QY	82	ProLysGlyProValThrGlnMetThrAsnValAspLysAspLeuValGlyTyrPheGln	101
DB	304	CCAAAGGGCCCAATACCCAGATGTACACTAATGTGGACAGACCTCTCGGCTGCAG	363
QY	102	AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu	121

DB	364	CGCGCCCGCGGGCGCGTCTCTTGCACACATGCACCTGTGGCAGCTCAGACCTTTACTTG	423
QY	122	ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu	141
DB	424	GTCACGAGACATGCTGACGTCATTCGGTGCAGCGGGGCGCACAGTAGGGGGAGCCTG	483
QY	142	LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro	161
DB	484	CTCTCCCCCAGGCTGCTCTTCTACTTGAAGGGCTCTTCGGGTGGTCCACTGCTCTCCCT	543
QY	162	AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla	181
DB	544	TCGGGGCAGCGTGTGGGCATCTTCGGGCTGCCGTATGCACCGGGGGTTCGAAAGCG	603
QY	182	ValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro	197
DB	604	GTGGACTTTGTGCCCGTAGAGTCCATGGAACCTACTATGCGGTCTCCG	651

Search completed: August 30, 2003, 19:48:21  
 Job time : 188.939 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:26:03 ; Search time 176.482 Seconds  
(without alignments)  
2560.981 Million cell updates/sec

Title: US-09-965-594-22

Perfect score: 1016

Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-MODEL=frame\_p2n.model -DEV-xlp  
-Q=/cgn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151920\_28367/app\_query.fasta\_1.2872  
-DB=Published\_Applications\_NA -GPMF=fastap -SUFFIX=tnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09965594.ecgn\_1.1.864.@runat\_29082003\_151920\_28367  
-NCPUP=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	100.0	594	10	US-09-965-594-23 Sequence 23, Appl

2	1010	99.4	594	10	US-09-965-594-21	Sequence 21, Appl
3	995	97.9	594	10	US-09-965-594-19	Sequence 19, Appl
4	980	96.5	594	10	US-09-965-594-17	Sequence 17, Appl
5	963	94.8	594	10	US-09-965-594-15	Sequence 15, Appl
6	936	92.1	594	10	US-09-965-594-25	Sequence 25, Appl
7	929	91.4	588	10	US-09-965-594-13	Sequence 13, Appl
8	902	88.8	588	10	US-09-965-594-4	Sequence 4, Appl
9	864.5	85.1	612	14	US-10-133-133A-6	Sequence 6, Appl
10	848.5	83.5	9646	9	US-09-742-659-3	Sequence 3, Appl
11	848.5	83.5	9646	10	US-09-238-076-1	Sequence 1, Appl
12	848.5	83.5	9646	11	US-09-995-937-1	Sequence 1, Appl
13	848.5	83.5	9646	11	US-09-917-563-1	Sequence 1, Appl
14	848.5	83.5	12980	10	US-09-238-076-5	Sequence 5, Appl
15	848.5	83.5	12980	11	US-09-995-937-5	Sequence 5, Appl
16	848.5	83.5	12980	11	US-09-917-563-5	Sequence 5, Appl
17	844.5	83.1	9379	9	US-09-916-359-1	Sequence 1, Appl
18	844.5	83.1	9416	10	US-09-238-076-19	Sequence 19, Appl
19	844.5	83.1	9416	11	US-09-995-937-19	Sequence 19, Appl
20	844.5	83.1	9416	11	US-09-917-563-19	Sequence 19, Appl
21	842	82.9	549	10	US-09-965-594-2	Sequence 2, Appl
22	842	82.9	2058	10	US-09-881-654-1	Sequence 1, Appl
23	842	82.9	2058	10	US-09-881-239-2	Sequence 2, Appl
24	841.5	82.8	836	10	US-09-921-397-120	Sequence 120, Appl
25	841.5	82.8	10803	10	US-09-747-419-17	Sequence 17, Appl
26	841.5	82.8	10803	14	US-10-259-275-17	Sequence 17, Appl
27	838.5	82.5	9416	10	US-09-929-955-13	Sequence 13, Appl
28	838.5	82.5	9416	13	US-10-104-966-13	Sequence 13, Appl
29	838	82.5	2061	10	US-09-929-955-16	Sequence 16, Appl
30	826.5	81.3	13910	11	US-09-919-901-1	Sequence 1, Appl
31	823.5	81.1	13910	11	US-09-919-901-8	Sequence 8, Appl
32	823.5	81.1	13910	11	US-09-919-901-15	Sequence 15, Appl
33	823	81.0	2064	11	US-09-884-456-69	Sequence 69, Appl
34	823	81.0	2523	11	US-09-884-456-85	Sequence 85, Appl
35	820.5	80.8	2073	14	US-10-133-133A-5	Sequence 5, Appl
36	820	80.7	6189	14	US-10-259-275-41	Sequence 41, Appl
37	820	80.7	7992	13	US-10-005-469-1	Sequence 1, Appl
38	820	80.7	7992	13	US-10-005-469-2	Sequence 2, Appl
39	820	80.7	7992	13	US-10-005-469-4	Sequence 4, Appl
40	820	80.7	7992	12	US-10-005-469-6	Sequence 6, Appl
41	820	80.7	8638	12	US-10-309-561-24	Sequence 24, Appl
42	820	80.7	8638	13	US-10-029-907-24	Sequence 24, Appl
43	820	80.7	8639	12	US-10-309-561-1	Sequence 1, Appl
44	820	80.7	8639	13	US-10-029-907-1	Sequence 1, Appl
45	820	80.7	8642	12	US-10-309-561-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-09-965-594-23  
; Sequence 23, Application US/09965594  
; Patent No. US20020106642A1  
; GENERAL INFORMATION:  
; APPLICANT: Wittekind, Michael  
; APPLICANT: Weinheimer, Steven  
; APPLICANT: Zhang, Yaqu  
; APPLICANT: Goldfarb, Valentina  
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for  
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies  
; FILE REFERENCE: DB17Sequences  
; CURRENT APPLICATION NUMBER: US/09/965,594  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/115,271  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Hepatitis C virus  
US-09-965-594-23

```
Alignment Scores:
Pred. No.: 4,57e-109 Length: 594
Score: 1016.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-23 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
DB 61 TACGCTCAGCAGCTGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAAACAGGTTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTCGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGGTGTTCTGTGGACCGTTTACCACGGTGTGTGTGTGTGTGTGTGT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGAAAGGTCGGTTACCCAGATGTACACCAAGTTGACAAAGACCTCGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCAGCGGTCCCGTTCCCTGACCCCGTCCACCTCGGTTCCCTCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGTCACGTCAGCTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCTCCCGTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTGGTATCTTCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTGACTTCATCCCGTTGAATCCCTGGAACACCATCGCTTCCCGG 591

RESULT 2
US-09-965-594-21
; Sequence 21, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 60/115,271
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus

US-09-965-594-21
Alignment Scores:
Pred. No.: 2,28e-108 Length: 594
Score: 1010.00 Matches: 196
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.41% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-21 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAAAAAAGGATCGGTTGTTATCGTCGGCGGTATCAACCTGTCGGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 40
DB 61 TACGCTCAGCAGCTGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 41 LysAsnGlnValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AAAAAACAGGTTGAAGTTCAGATCGTTTCCACCGCTACCCAGACCTTCCTCGCT 180
QY 61 ThrSerIleAsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTCCATCAACGGGTGTTCTGTGGACCGTTTACCACGGTGTGTGTGTGTGTGTGTGT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTrp 100
DB 241 TCCCGAAAGGTCGGTTACCCAGATGTACACCAAGTTGACAAAGACCTCGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CAGGCTCCAGCGGTCCCGTTCCCTGACCCCGTCCACCTCGGTTCCCTCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGTCACGTCAGCTTATCCCGGTTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerGlySerGlyProLeuLeuCys 160
DB 421 CTGCTGTCCTCCCGTCCGATCTCTACCTGAAAGGTTCCTCCGGTGGTCCGCTGCTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CCGGCTGGTCAGCTGTGGTATCTTCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTAAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTGTGACTTCATCCCGTTGAATCCCTGGAACACCATCGCTTCCCGG 591

RESULT 3
US-09-965-594-19
; Sequence 19, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: US/09/965,594
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 60/115,271
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
```





```
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-15

Alignment Scores:
Pred. No.: 6,71e-103 Length: 594
Score: 963.00 Matches: 187
Percent Similarity: 95.94% Conservativeness: 2
Best Local Similarity: 94.92% Mismatches: 8
Query Match: 94.78% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-15 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAATAAAAGGATCGTGTGTTATCGTGGCCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTTGAGGAGGTTGCCAAGAAACCTCCAGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AANAACCAAGTTGAGGTGAGTTGAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuThrPheValThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCCGCTGCTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGAAGGTGCCGTTATCCAGATGATACCAACGTTGACAAAGACCTGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CGGGCTCCGACGGTTCGGTTCCTGACCGCGTGCACCTCGGGTCTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGTCACGTCAGCTATCCCGGTTCGTCTGTTGGTACCTCCCGTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCTCTCCGTTGGTCCGCTGTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CGGGCTGGTCACGCTGTGGTATCTCCGCTGCTGCTGTTTCACCCGCGGTGGTGTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATCGCTTCCCG 591

RESULT 6
US-09-965-594-25
; Sequence 25, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/115,271
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-965-594-25

Alignment Scores:
Pred. No.: 9.3e-100 Length: 594
Score: 936.00 Matches: 185
Percent Similarity: 93.91% Conservativeness: 0
Best Local Similarity: 93.91% Mismatches: 12
Query Match: 92.13% Indels: 0
DB: 10 Gaps: 0

US-09-965-594-22 (1-197) x US-09-965-594-25 (1-594)
QY 1 MetLysLysLysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAla 20
DB 1 ATGAATAAAAGGATCGTGTGTTATCGTGGCCGTATCAACCTGTCGGTGACACCGCT 60
QY 21 TyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAsp 40
DB 61 TACGCTCAGCAGACTCGAGGTTGCTGGTGTGCATCATCACCTCCCTGACCGGTGCTGAC 120
QY 41 LysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAla 60
DB 121 AANAACCAAGTTGAGGTGAGTTGAGTTCAGATCGTTTCCACCGCTGCTCAGACCTTCTTGGCT 180
QY 61 ThrSerIleAsnGlyValLeuThrPheValThrValTyrHisGlyAlaGlyThrArgThrIleAla 80
DB 181 ACCTGCATCAACGGTGTGTTGCTGGACCGTTTACCACGGTGTGTTACCCGCTGCTACCATCGCT 240
QY 81 SerProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr 100
DB 241 TCCCGAAGGTGCCGTTATCCAGATGATACCAACGTTGACAAAGACCTGTTGGTTGG 300
QY 101 GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyr 120
DB 301 CGGGCTCCGACGGTTCGGTTCCTGACCGCGTGCACCTCGGGTCTCTCCGACCTGTAC 360
QY 121 LeuValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySer 140
DB 361 CTGGTTACCGTCACGTCAGCTATCCCGGTTCGTCTGTTGGTACCTCCCGTGGTTC 420
QY 141 LeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCys 160
DB 421 CTGCTGTCCCGCGTCCGATCTCTACCTGAAGGTTCTCTCCGTTGGTCCGCTGTGC 480
QY 161 ProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLys 180
DB 481 CGGGCTGGTCACGCTGTGGTATCTCCGCTGCTGCTGTTTCACCCGCGGTGGTGTAA 540
QY 181 AlaValAspPheIleProValGluSerLeuGluThrThrMetArgSerPro 197
DB 541 GCTGTTGACTTCATCCCGTTGAATCCCTGGAAACCAACCATCGCTTCCCG 591

RESULT 7
US-09-965-594-13
; Sequence 13, Application US/09965594
; Patent No. US20020106642A1
; GENERAL INFORMATION:
; APPLICANT: Wittekind, Michael
; APPLICANT: Weinheimer, Steven
; APPLICANT: Zhang, Yaqu
; APPLICANT: Goldfarb, Valentina
; TITLE OF INVENTION: Modified Forms of Hepatitis C NS3 Protease for
; TITLE OF INVENTION: Facilitating Inhibitor Screening and Structural Studies
; TITLE OF INVENTION: of Protease:Inhibitor Complexes
; FILE REFERENCE: DB17Sequences
; CURRENT APPLICATION NUMBER: 2001-09-27
```



## ; TITLE OF INVENTION: AND USE OF SAME TO TREAT VIRAL INFECTIONS

```
; FILE REFERENCE: NB 2021.00
; CURRENT APPLICATION NUMBER: US/10/133,133A
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: 60/286,983
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Hepatitis C. Virus
US-10-133-133A-6

Alignment Scores:
Pred. No.:      2,02e-91      Length: 612
Score:          864.50      Matches: 174
Percent Similarity: 90.77%      Conservative: 3
Best Local Similarity: 89.23%      Mismatches: 15
Query Match:      85.09%      Indels: 3
DB:              14      Gaps: 1

US-09-965-594-22 (1-197) x US-10-133-133A-6 (1-612)
```

```
QY 5 GlySerValValIleValGlyArqIleAsnLeuSerGlyAsp-----ThrAlaTyr 21
Db 19 GCTAGTGTGGTCAATTTGGGTAGGATCATTTTCCGGTAGTGGTATCATCGGGCTAC 78
QY 22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLys 41
Db 79 GCCCAGCAGACAGGGGCTCTAGGTGTCATATCACACGCTTAATCGCCGGGACAAA 138
QY 42 AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThr 61
Db 139 AACCAAGTGGAGGTGAGGTCCAGATTGTGTCACTGTCTCCCAAACTTCTCTGGCAACG 198
QY 62 SerIleAsnGlyValLeuTTPThrValTyrHisGlyAlaGlyThrArgThrIleAlaSer 81
Db 199 TGCATCAATGGGTGTCTGGACTGTCTACACGGGGCCGGACGAGGACCATCGCGTCA 258
QY 82 ProLysGlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTTPGln 101
Db 259 CCCAAGGTCTCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGCTGGCCC 318
QY 102 AlaProGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeu 121
Db 319 GCTTCGCAAGGTACCCGCTCAATGACACCCCTGCACTTGGGCTCCTCGGACCTTTACCTG 378
QY 122 ValThrArgHisAlaAspValIleProValArgArgGlyAspSerArgGlySerLeu 141
Db 379 GTCACGAGGCACGGGATGTCATCCCGTGGCGGGGGGTGATACGAGGGGACGCTG 438
QY 142 LeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysPro 161
Db 439 CTGCGCCCGCCGCGCATTTCTACTTGAAGGCTCTCTCGGGGGTCCGCTGTTGTGCCCC 498
QY 162 AlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyValAlaLysAla 181
Db 499 CGGGGGACCGCGTGGGCATATTTAGGGCCCGGTGTGCACCCGTGGAGTGGCTAAGCCG 558
QY 182 ValAspPheIleProValGluSerLeuGluThrThrMetArgSer 196
Db 559 GTGGACTTTATCCCTGTGGAGAACCTTAGAGAACCACTAGAGTGC 603
```

## RESULT 10

```
US-09-742-659-3
; Sequence 3, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
```

```
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 9646
; TYPE: DNA
; ORGANISM: Hepatitis C virus
US-09-742-659-3
```

```
Alignment Scores:
Pred. No.:      5,25e-88      Length: 9646
Score:          848.50      Matches: 168
Percent Similarity: 87.25%      Conservative: 10
Best Local Similarity: 82.35%      Mismatches: 17
Query Match:      83.51%      Indels: 9
DB:              9      Gaps: 1
```

```
US-09-965-594-22 (1-197) x US-09-742-659-3 (1-9646)
```

```
QY 3 LysLysGlySerValIleValGlyArgIleAsn----- 14
Db 335A COTAGGGCCAGGAGATACTGTTGGACCGCCGCAAGTGGTCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCGCCCATCACGGCTAGCCCGACGACGAGAGCCCTCCTAGAGTGTATAATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCACTCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGTCCAGATCGTGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTTPThrValTyrHisGly 73
Db 3534 GCTACCCAAACCTTCTCTGGCAACGTGCATCAATGGGGTATGCTGGACTGTCTACACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCGGAAACGAGGACCATCGATCACCCAGGCTCTGTCTCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTTPGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGCTGGCGCGCTCTCTCAAGGTTCCCGCTCATTCACACCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGGCTCTCTGGAGCTTTTACCTGGTTCAGAGGACGCGCGCATTTCCCGTGGCGCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTCGCCCGCGCCCATTTCTACTAAAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGTGCCCGGGGACACGCGGTGGGCTATTTCAGGGCGCGGTG 3893
QY 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3894 TGCACCCGTGGAGTGGCCAAAGCGGTGGACTTTATCTCTGTGGAGAACCTTAGACACACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCCC 3965
```

## RESULT 11

```
US-09-238-076-1
```



```

Pred. No.: 5,25e-88 Length: 9646
Score: 848.50 Matches: 168
Percent Similarity: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-995-937-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn-----14
Db 3354 CGTAGGGCCAGAGATACCTTGGCCAGCCAGCGAATGGTCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlnLys 33
Db 3414 TTGCTGGCCCATCAGCGGTACGCCAGCAGCAGAGGCTCTAGGTGTATATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTyrValIleHisGly 73
Db 3534 GCTACCAACCTTCTGGCAACGTGCATCAATGGGTATGCTATCCAGATGTATACCAATGTC 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCCGGAACGAGGACCATCGCATCCCAAGGTCTCTGTCATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTyrPheGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGTGGCCCTCTCAAGGTTCCTCAAGTTCCTGATTCACACCTGCACC 3713
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCGGACCTTACCTGGTACGAGGACGCCGATGTCNTCCGTGGCGCG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGTACCTGCTTCGCCCGCCCATTTCTTCTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGGCCCGGACACGCGTGGGCTTATTACGGCGCGGTG 3893

Alignment Scores: 5,25e-88 Length: 9646
Pred. No.: 848.50 Matches: 168
Score: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-917-563-1 (1-9646)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn-----14
Db 3354 CGTAGGGCCAGAGATACCTTGGCCAGCCAGCGAATGGTCTCCAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlnLys 33
Db 3414 TTGCTGGCCCATCAGCGGTACGCCAGCAGCAGAGGCTCTAGGTGTATATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuTyrValIleHisGly 73
Db 3534 GCTACCAACCTTCTGGCAACGTGCATCAATGGGTATGCTATCCAGATGTATACCAATGTC 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCCGGAACGAGGACCATCGCATCCCAAGGTCTCTGTCATCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTyrPheGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGTGGCCCTCTCAAGGTTCCTCAAGTTCCTGATTCACACCTGCACC 3713
QY 114 CysGlySerSerAspLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCGGACCTTACCTGGTACGAGGACGCCGATGTCNTCCGTGGCGCG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGTACCTGCTTCGCCCGCCCATTTCTTCTTGAAGGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTGTGGCCCGGACACGCGTGGGCTTATTACGGCGCGGTG 3893

RESULT 13
US-09-917-563-1
; Sequence 1, Application US/09917563
; Publication No. US20030073080A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
;
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

Qy 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
|||||  
Db 3894 TCACCCGTTGGAGTGGCTAAGCGGTGGACCTTATCCCTGTGGAGAACCTAGAGACAC 3953  
Qy 194 MetArgSerPro 197  
|||||  
Db 3954 ATGAGATCCCG 3965

RESULT 14  
US-09-238-076-5  
; Sequence 5, Application US/09238076  
; Patent No. US20020102540A1  
; GENERAL INFORMATION:  
; APPLICANT: RICE, CHARLES et al.  
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MO  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/238,076  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/034,756  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 6029-4831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 314-727-5188  
; TELEFAX: 314-727-6092  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12980 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-09-238-076-5

Alignment Scores:  
Pred. No.: 7,72e-88 Length: 12980  
Score: 848.50 Matches: 168  
Percent Similarity: 87.25% Conservative: 10  
Best Local Similarity: 82.35% Mismatches: 17  
Query Match: 83.51% Indels: 9  
DB: 10 Gaps: 1

US-09-965-594-22 (1-197) x US-09-238-076-5 (1-12980)

Qy 3 LysLysGlySerValIleValGlyArgIleAsn----- 14  
:|||||:|||||:|  
Db 3354 CGTAGGGCCGAGAGATGCTTGGCCAGCGCGAATGCTCCAGGGGTGGAGG 3413  
Qy 15 ---LeuSerGlyAspThrAlaIleGlnGlnThrArgGlyGluGlnGlnLys 33  
:|||||:|||||:|  
Db 3414 TTGCTGGCGCCCATCAGCGGTACGCCAGCAGACGAGCGCTCTAGGGTGTATATC 3473  
Qy 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53

Db 3474 ACCAGCCTGACTGGCGGGACAAAACCAAGTGGAGGTGAGGTCCAGATCGGTCAACT 3533  
Qy 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrThrValThrHisGly 73  
:|||||:|||||:|  
Db 3534 GCTACCCAAACCTTCTGGCAACGTGCATCAATGGGTATGCTGAGTCTCTACCAACGG 3593  
Qy 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyThrAsnVal 93  
:|||||:|||||:|  
Db 3594 GCCGGAACGAGACCATCGCATCACCACAGGTCTCTGTATCCAGATGTATACCAATGTG 3653  
Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113  
:|||||:|||||:|  
Db 3654 GACCAAGACCTTGGGCTGGCCGCTCTCAAGGTTCGCGCTCATTCACACCTGCACC 3713  
Qy 114 CysGlySerSerAspLeuTyLeuValThrArgHisAlaAspValIleProValArgArg 133  
:|||||:|||||:|  
Db 3714 TGGGCTCTCGGACCTTTACCTGGTCACGAGCAGCCGATGTCTATCCCGTGGCCCGG 3773  
Qy 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyLeuLysGlySer 153  
:|||||:|||||:|  
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCGCCCGGCCCATTTCTTACTTGAAGGCTCC 3833  
Qy 154 SerGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173  
:|||||:|||||:|  
Db 3834 TCGGGGGTCCGCTGTTGTCGCCCGCGGACACGCCGTGGCGCTATTCAGGGCGCGGTG 3893  
Qy 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193  
:|||||:|||||:|  
Db 3894 TGCACCGGTGGAGTGGCTAAGCGGTGGACTTATCCCTGTGGAGAACCTAGAGACAC 3953

194 MetArgSerPro 197  
|||||  
Db 3954 ATGAGATCCCG 3965

RESULT 15  
US-09-965-937-5  
; Sequence 5, Application US/09995937  
; Publication No. US20030028010A1  
; GENERAL INFORMATION:  
; APPLICANT: RICE, CHARLES et al.  
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.  
; STREET: 7733 FORSYTH BLVD., SUITE 1400  
; CITY: ST. LOUIS  
; STATE: MO  
; COUNTRY: USA  
; ZIP: 63105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/995,937  
; FILING DATE: 28-NO. US20030028010A1-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/034,756  
; FILING DATE: 04-May-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HOLLAND, DONALD R.  
; REGISTRATION NUMBER: 35,197  
; REFERENCE/DOCKET NUMBER: 6029-4831  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 314-727-5188  
; TELEFAX: 314-727-6092  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12980 base pairs

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-937-5

Alignment Scores:
Pred. No.: 7.72e-88 Length: 12980
Score: 848.50 Matches: 168
Percent Similarity: 87.25% Conservative: 10
Best Local Similarity: 82.35% Mismatches: 17
Query Match: 83.51% Indels: 9
DB: 11 Gaps: 1

US-09-965-594-22 (1-197) x US-09-995-937-5 (1-12980)

QY 3 LysLysGlySerValValIleValGlyArgIleAsn----- 14
Db 3354 CgtAGGGCCAGGAGACTGCTTGGCCACCGCGGAATGGTCTCCAAGGGGTGGAGG 3413
QY 15 ---LeuSerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLys 33
Db 3414 TTGCTGGCGCCCATCAGCGCTACGCCAGCAGCAGAGAGCGCTCTTAGGGTGTATAATC 3473
QY 34 ThrSerHisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThr 53
Db 3474 ACCAGCTGACTGGCCGGGCAAAACCAAGTGGAGGGTGAGGTCCAGATCGTGTCAACT 3533
QY 54 AlaThrGlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrThrValThrHisGly 73
Db 3534 GCTACCCAAACCTTCCTGGCAACGTGCATCAATGGGTATGCTGGACTGTCTACCACGGG 3593
QY 74 AlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyrThrAsnVal 93
Db 3594 GCCGAAGCAGGACCATCGCATCACCCCAAGGGTCTGTCTCCAGATGTATACCAATGTG 3653
QY 94 AspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113
Db 3654 GACCAAGACCTTGTGGGCTGGCCGCTCTCCTCAAGGTTCCCGCTCATTTGACACCCCTGCACC 3713
QY 114 CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArg 133
Db 3714 TCGGCTCTCTCGGACCTTTACTGTGTACAGGACGACGCGGATGTCATTCCGTCGCCGG 3773
QY 134 ArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySer 153
Db 3774 CGAGGTGATAGCAGGGGTAGCTGCTTTCCGCCCGGCCCATTTCTTCTACTTGAAGAGCTCC 3833
QY 154 SerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal 173
Db 3834 TCGGGGGGTCCGCTGTTGTCGCCCGGCGGACACCGCTGGGCTATTTCAGGGCCGCGGTG 3893
QY 174 SerThrArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThr 193
Db 3894 TGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAACCTAGAGACAACC 3953
QY 194 MetArgSerPro 197
Db 3954 ATGAGATCCCCG 3965
```

Search completed: August 31, 2003, 04:55:02  
Job time : 189.482 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 1910.31 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-22  
Perfect score: 1016  
Sequence: 1 MKKGSVVIVGRINLSGDTA.....VAKAVDFIPVESLETTMRSP 197  
Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Command line parameters: -DEV-xlp  
-Q/cgn2\_1/USPTO\_spool/US09965594/runat\_29082003\_151919\_28322/app\_query.fasta\_1.2872  
-DB-EST -QFMT-fastap -SUFFIX-rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=0 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE-LOCAL  
-OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN\_1\_1\_12630 -runat\_29082003\_151919\_28322 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104.5	10.3	1031	14	CB950999
2	101	9.9	1199	13	CB950999
C 3	100.5	9.9	1403	13	BO926101
C 4	98	9.6	644	29	BX238988
C 5	97.5	9.6	629	10	BO989727
C 6	96	9.4	1146	12	BM915803
C 7	95.5	9.4	701	10	BF863244
C 8	95.5	9.4	772	29	CC406704
C 9	95.5	9.4	789	29	CC406705
C 10	95.5	9.4	984	10	BF304699
C 11	95	9.4	701	29	B2342381
C 12	94.5	9.3	1082	29	CNS0608H
C 13	94.5	9.3	1440	12	BM467279
C 14	93	9.2	528	12	BM402566
C 15	93	9.2	560	28	AQ538021
C 16	93	9.2	1213	13	B0541777
C 17	92.5	9.1	938	13	B0894657
C 18	92.5	9.1	1026	29	B2567288
C 19	92	9.1	617	10	BE055938
C 20	92	9.1	886	9	AL571605
C 21	91.5	9.0	528	28	AQ620249
C 22	91.5	9.0	958	10	BG420860
C 23	91.5	9.0	1035	10	BE888775
C 24	91.5	9.0	1733	12	BM553374
C 25	91	9.0	580	14	CA728398
C 26	91	9.0	736	12	B1459445
C 27	91	9.0	733	13	B0402910
C 28	91	9.0	866	13	B0219343
C 29	91	9.0	906	13	BX434207
C 30	91	9.0	917	12	B1911168
C 31	90.5	8.9	586	12	B1329116
C 32	90.5	8.9	670	29	B2552327
C 33	90.5	8.9	812	13	B0299264
C 34	90.5	8.9	814	11	CNS09179
C 35	90.5	8.9	817	13	B0240438
C 36	90.5	8.9	824	13	B0396924
C 37	90.5	8.9	878	13	B0365755
C 38	90.5	8.9	920	13	B0593458
C 39	90.5	8.9	1106	13	B0956626
C 40	90.5	8.9	1141	11	AK080345
C 41	90	8.9	500	12	BM708007
C 42	90	8.9	569	12	BM825317
C 43	90	8.9	590	10	BE382750
C 44	90	8.9	631	10	AW961059
C 45	90	8.9	658	12	BM830847

ALIGNMENTS

RESULT 1  
CB950999  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

CB950999 1031 bp mRNA linear EST 29-APR-2003  
AGENCOURT\_13445496 NIH\_MGC\_177 Mus musculus CDNA clone  
IMAGE:30316162 5', mRNA sequence.  
CB950999  
CB950999.1 GI:30205777  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Muridae; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 1031)

# AUTHORS TITLE JOURNAL COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Michael Brownstein Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDCM107 row: b column: 11  
High quality sequence stop: 333.  
Location/Qualifiers

## SOURCE

1. 1031  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30316162"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_lib="NIH\_MGC\_177"  
/note="Organ: liver; Vector: pDNR-LIB; Site\_1: SfiI  
(ggccattatgcc); Site\_2: SfiI (ggccattcgcc); cDNA made  
by oligo-dT priming and directionally cloned. 5' and 3'  
adaptors were used in cloning as follows:  
5'-AAGCAGTGGTATCAGCAGAGTGGCCATACGGCGGG-3' and  
5'-ATTCTAGACCGGCGGCACATG-DT(30)NN-3'. Full-length  
enriched library was constructed using the Clontech  
Creator SMART kit and size-selected to contain the 0.5 kb  
size fraction. Library created in the laboratory of M.  
Brownstein (NIH, NIH). Note: this is a NIH-MGC Library."

BASE COUNT  
ORIGIN

235 a 309 c 211 g 275 t 1 others  
Alignment Scores:  
Pred. No.: 3.53 Length: 1031  
Score: 104.50 Matches: 51  
Percent Similarity: 41.10% Conservative: 16  
Best Local Similarity: 31.29% Mismatches: 62  
Query Match: 10.29% Indels: 35  
DB: 14 Gaps: 8

US-09-965-594-22 (1-197) x CB950999 (1-1031)

Qy 44 ValGluGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIle 63  
Db 395 ATTCAGGCTATCCCAAAACAAGAGATACATCGGCAGCTTTTCCT---CACTCATTT 451  
Qy 64 AsnGlyValLeuThrValThrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83  
Db 452 TTGGGCACACTGGTCCGTGGGCACAT-----ATCATCGCCCTAA 493  
Qy 84 GlyProValThrGlnMetTyrThrAsnValAspLysValLeuValTyrGlnAlaPro 103  
Db 494 GGGCCTTTCACAAA-----ACACTTAACCT-CCTTGCCTGGCCTGGCATGTGGG 543  
Qy 104 Gln-----GlySerArgSerLeuThrProCysThrCysGlySerSerAsp 118  
Db 544 CAAAGAGACGGTTTGGCGCTCTTGGCGCCCTTGGCGCCCTTGGAGACCATGGCG 603  
Qy 119 LeuTyrLeuValThrArgHisAlaAsp-ValIleProValArgArgGlyAspSerAr 138  
Db 604 -----ACCACCATGGGCTGTGTGTTCCCGCCTCTCCCGTGGGCAATACA 651  
Qy 138 gGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGly----- 155  
Db 652 AAACNCCCTTAACCGTCCCTCCCAACAATATTCTTCAAGCGTCTCTGATTTCCCTAA 711  
Qy 156 -GlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerth 175  
Db 712 GTCCCCCTTTGTTACCCAGACCATTTGTGGGACACAGCGCTCTTTTATCTTC 771

Qy 175 rArgGlyValAlaLysAlaValAspPheIleProValGluSerLeuGluThrThrMetAr 195  
Db 772 C-----CCCCCTCATGCTCTT---CCCACACGGCG 798

Qy 195 gSerPro 197  
Db 799 AACACCC 805

## RESULT 2

BO892487  
LOCUS BO892487 1199 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGENCOURT\_8417538 Lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
IMAGE:6192708 5', mRNA sequence.

## ACCESSION

BO892487

## VERSION

BO892487.1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 1199)

## AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

## TITLE

Unpublished

## JOURNAL

Contact: Robert Strausberg, Ph.D.

## COMMENT

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM13595 row: c column: 13  
High quality sequence start: 57  
High quality sequence stop: 394.  
Location/Qualifiers

## FEATURES

Source

1..1199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6192708"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCACGGTCCG-3' and  
5'-GACTAGTCTCTAGTCGGAGCGCGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life  
Technologies."

BASE COUNT 255 a 362 c 343 g 211 t 28 others  
ORIGIN

## Alignment Scores:

Pred. No.: 9.59 Length: 1199  
Score: 101.00 Matches: 50  
Percent Similarity: 34.74% Conservative: 24  
Best Local Similarity: 23.47% Mismatches: 80  
Query Match: 9.94% Indels: 59  
DB: 13 Gaps: 9

US-09-965-594-22 (1-197) x BO892487 (1-1199)

Qy 16 SerGlyAspThrAlaTyrAlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSer 35  
Db 337 GCAGGAGAGAAACCTTACCCCAACAG-----AAGGCA 369

```

Qy 36 HisThrGlyArgAspLysAsnGlnValGluGlyValGlnIleValSerThr----- 53
Db 370 CATGGGGAATCCCGCTTCACAGAGAGGCTCTTCATGTTTCTGAAACATAACCG 429
Qy 54 AlaThrGlnThrPheLeu-----AlaThrSerIleAsnGlyValLeuTrpThr 69
Db 430 CCAGCCACTGCTTCATGTAATGACCTTCCACACACACACAGGGCAGCATGGGAI 489
Qy 70 ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValThrGlnMet 89
Db 490 CCATTTTAAAGAGGTGCTCTTAAATCATGCGCCACGCGCGCTGATCTTCCA 549
Qy 90 TyrThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeu 109
Db 550 TTTACCACATGTGACAGTGACTTT-----CysGlySerSerAspLeuTyr 120
Qy 110 ThrProCysThr-----CysGlySerSerAspLeuTyr 120
Db 577 GCTGCTGCACAGCACCCCATGACCATGTGGGCTTATGTGGAACGCGGAGGCTTC 636
Qy 121 -LeuValThr-----ArgHisAlaAspValIleProValArg----- 132
Db 637 ATGGCCACTCCCTCTATATAAACACAGCCCAAGCTGTTCCATGGCGCGGCTGGTGT 696
Qy 133 -----ArgArgGlyAspSerArgGlySerLeuLeu----- 142
Db 697 TTGCAGCGCAAGCGGGTGGGGCATGTAGTACTCGGGGCGGATCTCTGAAACACC 756
Qy 143 -SerProArgProIleSerTyrLeuLys-----GlySerSerGlyGlyPr 157
Db 757 CCACCTCGGCCCCACCATGCGGTAAAGCTCCCTTTACAGCCACCGCGCGGCCCC 816
Qy 157 oLeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgG1 177
Db 817 CCTAACATCTCTACCTCCGCGCGCGGGGAGAGCTGGGGCATACGGGCTCAGG 876
Qy 177 yValAlaLysAlaValAspPheIleProValGluSer 189
Db 877 CGTTTAAAGCCCGCGCTTCGCGCGCGGGAACGA 913

RESULT 3
BQ926101/c
LOCUS BQ926101 1403 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8752655 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6335718
5', mRNA sequence.
ACCESSION BQ926101
VERSION BQ926101.1 GI:22341132
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution by: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAH13798 row: j column: 07
High quality sequence stop: 101.
Location/Qualifiers
1. 1403
/mol_type="mRNA"
/organism="Mus musculus"
source

```

```

/db_xref="taxon:10090"
/clone="IMAGE:6335718"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_130"
/note="Organ: oocytes; Vector: pCMV-SPORT6.1.cdbb;
Site_1: EcorV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 297 a 521 c 237 g 345 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 13.2 Length: 1403
Score: 100.50 Matches: 56
Percent Similarity: 35.32% Conservative: 15
Best Local Similarity: 27.86% Mismatches: 74
Query Match: 9.89% Indels: 56
DB: 13 Gaps: 9
US-09-965-594-22 (1-197) x BQ926101 (1-1403)
Qy 4 LysGlySerValValIleValGlyArgIleAsnLeuSerGlyAspThrAlaTyrAlaGln 23
Db 1381 AGAGGGTGTTCACGCGTCAGGACAGGTC---GCCGCACATCGACGGTCGGCCAGAG 1325
Qy 24 GlnThrArgGlyGluGlnGlyThrGlnLysThrSerHisThrGlyArgAspLysAsnGln 43
Db 1324 ACTTGTGGGGCGCGCTTGGCGCATACCCGGGTCGGATCGAGGTCAGGCGCTTGAT 1265
Qy 44 ValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIle 63
Db 1264 ACAGAGGGGAAA----- 1253
Qy 64 AsnGlyValLeuTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys 83
Db 1252 CAGGGGGA---TGGTTATCACGGGCTGGGCGAGGTACT-----TCCCTAAA 1208
Qy 84 --GlyProValThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaP 103
Db 1207 GCGGCGGCTGGCGGAGTATATATACCGCGAGTGGCAAGCGCGGCGGTGGAACGTTG 1148
Qy 103 roGlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuVal 123
Db 1147 ACCAA---CAGAGGCACTTCAGCCCTCCCTCGTGGGCTGTCGATAATAACAAATGTG 1091
Qy 123 hrArgHisAlaAspValIleProValArgArgGlyAsp----- 136
Db 1090 CAGGCGCAGGTGATGTGTTACTACCGCGAGCCGCTCCACGCGGCTCTCTACAGA 1031
Qy 137 -----SerArgGlySerLeuLeuSerProArgProIle-SerTyrLeuLysGlySer 153
Db 1030 CGGCCCGCTCCCGCGCAAC-----AGCGTAATAATCATATCGGCGCGGAT 983
Qy 154 SerGly-----GlyProLeuLeu 159
Db 982 TTGCGATTCGCGGAGAGCGCGGCTGCGGGGCGCGCGCTCGCGGCTGAGG 923
Qy 160 CysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGlyVal 178
Db 922 CGCAGGAGAGGC-----GCCGTGTTTCGCGGCTGAGGACGAGCGCGGCTG 875

RESULT 4
BX238988/c
LOCUS BX238988 644 bp DNA linear GSS 29-JAN-2003
DEFINITION Danio rerio genomic clone DREY-283113, genomic survey sequence.
ACCESSION BX238988
VERSION BX238988.1 GI:28161322
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

```

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	BASE COUNT	ORIGIN
1	(bases 1 to 644)	Humphray, S.J., Huckle, E. and Durham, J.L.	Direct Submission	Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk unpublished This sequence was generated from the T7 end of BAC 283L13. 283L13 is part of the Baniokey BAC library created by R. Piasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/projects/D-rerio/.	<p>sequence.</p> <p>CG089727</p> <p>EST.</p> <p>Mus musculus (house mouse)</p> <p>ORGANISM</p> <p>Mus musculus</p> <p>Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.</p> <p>1 (bases 1 to 629)</p> <p>NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a>.</p> <p>National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index</p> <p>Unpublished</p> <p>Contact: Robert Strausberg, Ph.D.</p> <p>Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a></p> <p>Tissue Procurement: David Segal Ph.D., Herbert Morse M.D.</p> <p>cDNA Library Preparation: Life Technologies, Inc.</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)</p> <p>DNA Sequencing by: Washington University Genome Sequencing Center</p> <p>Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a></p> <p>MGI:1477610</p> <p>Seq primer: -40UP from Gibco</p> <p>High quality sequence stop: 422.</p> <p>Location/Qualifiers</p> <p>1..629</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:3977578"</p> <p>/issue_type="NK cells (flow-sorted)"</p> <p>/lab_host="DH10B (T1-resistant)"</p> <p>/clone_lib="NCI-CGAP Sp2"</p> <p>/note="Organ: spleen; Vector: pCMV-SPORT6 (Life Technologies); mRNA made from flow-sorted NK cells, cDNA made by oligo-dT priming. Directionally cloned. Average insert size 1.5 kb. Primary library, non-amplified. cDNA Library Preparation: David B. Krizman, Ph.D."</p> <p>BASE COUNT 131 a 156 c 150 g 191 t</p> <p>ORIGIN</p> <p>Alignment Scores:</p> <p>Pred. No.: 9.19 Length: 629</p> <p>Score: 97.50 Matches: 48</p> <p>Percent Similarity: 39.23% Conservative: 23</p> <p>Best Local Similarity: 26.52% Mismatches: 59</p> <p>Query Match: 9.60% Indels: 51</p> <p>DB: 10 Gaps: 12</p> <p>US-09-965-594-22 (1-197) x BG089727 (1-629)</p> <p>Qy 36 HisThrGlyArgAspLysAsnGlnValGluGlyValGlnValSerThrAlaThr 55</p> <p>Db 620 CATCCGGGTAAAG-----GAAGGAGACACATCATCCTTGTGCA--- 579</p> <p>Qy 56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuThrValTyHisGly----- 73</p> <p>Db 578 AAAACATTTCCCATTCACATATAAATGCT-----ACTATCTTCAGCTGACATC 528</p> <p>Qy 74 -----AlaGlyThrArgThrIleAlaSer 81</p> <p>Db 527 ATGCTGTTAAAGCTGGAGTAGCCAGAGACTAAAGCTGTGAGACCCCTCAAGTTG 468</p> <p>Qy 82 ProLysGlyProValThrGlnMetTyThrAsnValAspLysAspLeuValGlyTrpGln 101</p> <p>Db 467 CCCAGATCCCAATCCCGGGTGAAGCAGNCAATGTC---TGCAGTGTGGCTGGCTGG--- 414</p> <p>Qy 102 AlaProGlnGlySerArgSerLeu-----ThrProCysThrCysGlySerSerAspLeu 119</p> <p>Db 413 -----GGGTCAAGGTCCTCAATGACATTAAGACATCTGCCCGCTCGGAGGTT 363</p>		

Query Match:	9.45%	Indels:	57
DB:	12	Gaps:	10
 US-09-965-594-22 (1-197) x BM915803 (1-1146)			
Qy	22 AlaGlnGlnThrArgGlyGluGlnGlyThrGlnLysThrSer-HisThrGlyArgAspLy 41 		
Dd	1098 GCGCAGGCGGTCTGCTGGCAGCGAGGTCGTCTCCGCCTCCTCCTACGTCCTCGTGGAG 1039 		
Qy	41 sASnGlnValGlucylCluValGlnIleValSerThrAlaThrGlnThrPheLeuAlaTh 61 :::     :: : :::: :		
Dd	1038 CTGAGGGACAGAGGT-----CTACGCCGTGGGTAGGGGA 1003 :::     :::		
Qy	61 rserIleAsnGlyValLeuTrpThrValTyHis-----GlyAlaGlyThr-- 76 :::    :::		
Dd	1002 CGCGCGTGTTGGATGTTGG-----PATCACTCCCGCGCGGGGGAGGTACGTG 949 :::    :::		
Qy	77 -----ArgThrIleAlaSerPro-----LysGlyProValThrGlnMetTy 90     :: : :::: :		
Dd	948 AGCGAGGGCGCGCGTCTCGGGCGCGCGCGCGGGCGCG----- 903     :::    :::		
Qy	90 rThrAsnValAspLysAspLeuValGlyTrpGlnAlaProGln-----GlySerAr 107 :::    :::		
Dd	902 -----CAGATGTGCGGTGGGAAGCGCCGCTGCGCGCGGTGGGGCGCAG 859 :::    :::		
Qy	107 gserLeuThrProCysThrCysGlySerSerAspLeuTyIleuValThrArgHisAlaAs 127 		
Dd	858 ACITGCTTGTGCTTTCTGTGG----- 834 		
Qy	127 pValIleProValIArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIl 147 		
Dd	833 -----CGGAGGCGCGCGCGCGCGCGCTAGTGTGGCGCGGTCCCCCTCT 787 		
Qy	147 eSerTyLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGl 167 :      :::		
Dd	786 CCGGTATCTACGGCGCGCGAGGACCACATCTCTCTCCG-----TG 742 		
Qy	167 yIlePheArgAlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---ilePr 186          :::		
Dd	741 GGCTTCCGGCGTCTGTGTCTTCGCGGTCTGCGCGGGGGGGGGGTTTCGCGTACC 682 		
Qy	186 oVal 187 		
Dd	681 TTG 678 		
 RESULT 7			
BF863244			
LOCUS	701 bp mRNA linear EST 19-JAN-2001		
DEFINITION	963042C02.xl C. reinhardtii CC-1690, Stress condition I, normalized Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.		
ACCESSION	BF863244		
VERSION	BF863244.1 GI:12253388		
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii		
ORGANISM	Chlamydomonas reinhardtii		
REFERENCE	Eukaryota: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.		
AUTHORS	Grossman,A., Davies,J.J., Federspiel,N., Harris,E., Hauser,C., Lefebvre,P., Mcdermott,J.P., Shrager,J., Silflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 3		
JOURNAL	Unpublished		
COMMENT	Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu.		
FEATURES	Location/Qualifiers 1..701		

[illegible]

```

Qy 178 ValAlaLysAlaValAspPhe-----ilePro 186
    |||
Db 244 -----CCATTGCATCAACAGGAGAGCTTCACCATTCAGGCATATGCAATTCCC 194

Qy 187 ValGluSerLeuGlu 191
    |||
Db 193 TTGGATACATTGGAG 179

RESULT 9
LOCUS CC406705 789 bp DNA linear GSS 19-MAY-2003
DEFINITION PUHKL2TD 2M_0.6_1.0_KB Zea mays genomic clone ZMBBTA469B24,
            genomic survey sequence.
ACCESSION CC406705
VERSION CC406705.1 GI:30886795
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
TITLE 1 (bases 1 to 789)
JOURNAL Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
COMMENT A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
            Maize Genomics Consortium
            Unpublished
            Other GSSs: PUHKL2TB
            Contact: Cathy Whitelaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: Tf
Class: sheared ends.
FEATURES             Location/Qualifiers
    source            1..789
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMBBTA469B24"
                     /clone_lib="2M_0.6_1.0_KB"
                     /note="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
BASE COUNT          210 a 220 c 203 g 156 t
ORIGIN
Alignment Scores:   19.6          Length: 789
Pred. No.:          95.50         Matches: 49
Score:              39.46%        Conservative: 24
Percent Similarity: 26.49%        Mismatches: 67
Best Local Similarity: 9.40%       Indels: 46
Query Match:        29            Gaps: 10
DB:

US-09-965-594-22 (1-197) x CC406705 (1-789)

Qy 29 GlnGlyThrGlnLysThrSerHisThrGlyArgAspLysAsnGlnVal---GluGlyGlu 47
    |||
Db 110 CAGTCTCGTTCAGGCGCTCACCCCGCGAGATGGGAGACGAGTAAACAGGGCTAT 169

Qy 48 ValGlnIleValSerThrAlaThrGlnThrPheLeuAlaThrSerIleAsnGlyVal--- 66
    |||
Db 170 GTTACATTTGTGACG-----AGCCCTATGTACGAGGCCACCGCTGTCGCCAGGCTATTCT 223

Qy 67 LeuThrThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProVal 86
    |||
Db 224 ATCTGGAGG-----TCACAGACTTTGCTGACGATGATGATCAAGGGTG 265

Qy 87 ThrGlnMetTyrThrAsnValAspLysAspLeuValGlyTyr-----GlnAlaPro 103
    |||
Db 266 ACACAGATGATGACCCAGG-----AAGACGAGGAGCCCGAGTGGTTCCCTACATCAATAA 319

```

```

Qy 104 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeu----- 119
    |||
Db 320 CTGGGATTAGAGGAGGAGCACCACCATGCGAGCTCGGGTAGTCTCCTCAAGGTCAGGAGCTGC 379

Qy 120 -----TyrLeuValThrArgHisAlaAspValIleProValArgArgGlyAspSer 137
    |||
Db 380 TGGCCCTACTTGACACAGGGTTCAACACATACACTTCATCACTGCAGGCGCA-CAACAG 438

Qy 138 ArgGlySerLeuLeuSerProArgProIleSerTyrLeuLysGlySerSerGlyGlyPro 157
    |||
Db 439 CTGGGGTTACTTTGAACCCACACAGGTCGCGCATGTCAAGGTGGCAATGGAGACCCA 498

Qy 158 LeuLeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaValSerThrArgGly 177
    |||
Db 499 GTTTTCTGCCAG-----GGAGTAACTCGTCGCGCA----- 528

Qy 178 ValAlaLysAlaValAspPhe-----IlePro 186
    |||
Db 529 -----GCCATTGCATCAACAGGAGAGTTCACCATTCAGGCATATGCAATTCCC 579

Qy 187 ValGluSerLeuGlu 191
    |||
Db 580 TTGGATACATTGGAG 594

RESULT 10
LOCUS BF304699/c 984 bp mRNA linear EST 21-NOV-2000
DEFINITION BF304699
            mRNA sequence.
ACCESSION BF304699
VERSION BF304699.1 GI:11251586
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-research.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCMI005 row: g column: 13
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
    source            1..984
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4122276"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; CDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCAGGAG(G). Size-selected
            for average insert size 1.8kb. Library constructed by
            Ling Hong in the laboratory of Gerald M. Rubin (University
            of California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).".
BASE COUNT          133 a 329 c 351 g 171 t
ORIGIN
Alignment Scores:   26.1          Length: 984
Pred. No.:

```

end repaired, adaptor ligated and size fractionated using  
sephadex. The resulting fragments were between 0.8 and 3  
kb and were cloned into the vector (.x/y reads in M13mp19,  
.b/g reads in pUC19). The same ligation was transformed in  
either JM107 or DH5a."

BASE COUNT 108 a 251 c 232 g 110 t

ORIGIN

Alignment Scores:

		18.8	Length:	701
	Pred. No.:	95.00	Matches:	61
	Score:	36.11%	Conservative:	17
	Percent Similarity:	28.24%	Mismatches:	70
	Best Local Similarity:	9.35%	Indels:	69
	Query Match:	29	Gaps:	11
	DB:			

US-09-965-594-22 (1-197) x B2342381 (1-701)

QY	23	GlnGlnThrArgGlyGluGlnGlyThrGln----	LysThrSerHisThrGlyArgAspLys	41
DB	619	CGCGCCCGGAGCGCCGCAAAACAACACTCAAACTTGCTCAACCGCGC-----	569	
QY	42	AsnGlnValGluGlyGluValGlnIleValSerThrAlaThrGlnThrPheLeu-----	59	
DB	568	-----GTGCGTGTGCAGCGGAGCTTGGACGCTGCTGCTCGCGC-CTCGCGGCC	532	
QY	60	-----AlaThrSerIleAsnGlyValLeuTyrThrValTyrHis	72	
DB	521	TGGGCGCACCACCGCCACCGCCCATCGGTGCGGGGTGTTTTTTCACGACTACCGCAC	462	
QY	73	GlyAlaGlyThrArgThr-----IleAlaSerProLysGlyProValThrGlnMet	89	
DB	461	GC CGCGGGAGGACGACGAGCGTGGCGCTGGCGCGCGCGCGCGCGCGTGGGAATTA	402	
QY	90	TyrThrAsnValAspLysAspLeuValGlyTyrGlnAlaProGlnGlySerArg---Ser	108	
DB	401	TGGCGCGCAATAACTCGGCACCTACCGGAGCGCGGACGAGAGTTCGCGCTGCTCT	342	
QY	109	LeuthrProCysThr-----	113	
DB	341	CTCCCTCCCTCTGCTGTCTCCACCGAGTGCCCGCGCGCGCTACGGGGACAGGGA	282	
QY	114	-----CysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProVal	131	
DB	281	AACTGGTGTGTGAGGATCACGTGGAGGGCGGACGACGACGA-----	237	
QY	132	ArgATGArgGlyAsp-----SerArgGlySerLeuLeu	142	
DB	236	CGACGACGGGTGGAGACCCGGTTTGACGGCGCGCGCGCGCGCGCGGCGCACCTCTGC	177	
QY	143	SerProArgProIleSerTyrLeuLysGly-----Ser	153	
DB	176	TCGTCTCGCGCGTGGGAATATTTACTGTGCGCTCGCGCCCGCGCGCGCGCTCGGCCG	117	
QY	154	SerGlyGlyProLeu-LeuCysProAlaGlyHisAlaValGlyIlePheArgAlaAlaVal	173	
DB	116	CGGGGGGAGCGGTGGAGATGCGTGGCAGTGGCAGCGCGCT-----GCCTAGCT	69	
QY	173	lSerThrArgGlyVal-----AlaTysAlaValAspPheIlePro	186	
DB	68	GAGTCACCGCGGAGTCATCAGTGCCTCGCTTGTGGCTGCCT	23	

RESULT 12

CNS06QHN	LOCUS	CNS06QHN	1062 bp	DNA	linear	GSS 05-JUL-2001
DEFINITION	T3 end of clone AW0AA00603 of library AW0AA from strain CLIB 89 of					
ACCESSION	Yarrowia lipolytica, genomic survey sequence.					
VERSION	AL410673					
KEYWORDS	AL410673.1 GI:12179275					
SOURCE	GSS.					
ORGANISM	Yarrowia lipolytica					
	Yarrowia lipolytica					
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					





full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

BASE COUNT 330 a 442 c 308 g 355 t 5 others  
ORIGIN

## Alignment Scores:

Pred. No.: 54.2 Length: 1440  
Score: 94.50 Matches: 54  
Percent Similarity: 36.89% Conservative: 22  
Best Local Similarity: 26.21% Mismatches: 95  
Query Match: 9.30% Indels: 35  
DB: 12 Gaps: 9

US-09-965-594-22 (1-197) x BM467279 (1-1440)

Qy 12 AtgIleAsnSerGlyAspThrAlaTyAlaGlnGlnThrArgGlyClnGlnGlyThr 31  
Db 753 CAATCACCATACTCCGAGATGTTCTCTCTGT-----TTAAGCGTCACACA 800  
Qy 32 GlnLysThrSerHisThrGlyArgAspLysAsnGln-----ValGlu-GlyGluValG1 49  
Db 801 CCGCGCACCAACTCTACTCCGCCACACACAAATATATCTTCTGGAGCGGAATATCTT 860  
Qy 49 nIleValSerThrAlaThrGlnThrPheLeu---AlaThrSerIleAsnGlyValLeuTr 68  
Db 861 CTTCCGCCCCAGAGCAAGATGTTCTTTTATTTAGAAACGGGAGGGGCTTC 920  
Qy 68 pThrValThrHisGlyAlaGlyThrArgThrIleAlaSerProLysGly-----Pr 85  
Db 921 ATTTTTTTCCTCGCAGGCGACCTCTCGAATCCCGAGCGCGCGTTCCTCGCTCC 980  
Qy 85 oValThrGlnMetTyThrAsnValAspLysAspLeuValGlyTrp----- 100  
Db 981 TACATCCCGAGTATATAAT-----CCCGGTGGGTGGGACGTTTCT 1022  
Qy 101 -----GlnAlaProGlnGlySerArgSerLeuThrProCysThrCysGlySer-- 116  
Db 1023 ACACCCACACCGTGGCCCTCTCTATCTATCTGCTATCTATCTCCATCTGTGTGCTCC 1082  
Qy 117 -----SerAspLeuTyLeuValThrArgHisAlaAspValIleProVa 131  
Db 1083 CCACACCGCATTTTACTCCCGCTATATCTTCTGNTGCGGCGGAGCGCCCGC 1142  
Qy 131 lArgArgArgGlyAspSer-----ArgGlySerLeuLeuSerProArgProIleSerTy 149  
Db 1143 TAGGGGTGGGGGGGGCCATTTTTCACGGTAGC---ACATCGCCCGCCCTCATTTT 1199  
Qy 149 rLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePh 169  
Db 1200 TTGGTGGGGGGGGGGCGGCCACCCCTCACCCCTCGTGGGGGAGTCTGCTCTCTC 1259  
Qy 169 eArg-AlaAlaValSerThrArgGlyValAlaLysAlaValAspPheIleProValGluS 189  
Db 1260 CTCTAGCACAACCTTCATGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1319  
Qy 189 erLeuGluThrThr 193  
Db 1320 TTCTCGTCAACACA 1333

## RESULT 14

BM402566

LOCUS

DEFINITION BM402566 528 bp mRNA linear EST 01-JUL-2002  
resurrection plant Selaginella lepidophylla (EST) collection from the  
lepidophylla cDNA clone SLA005F12 5, mRNA sequence.

## ACCESSION

VERSION

BM402566

KEYWORDS

SOURCE

ORGANISM

Selaginella lepidophylla  
Selaginella lepidophylla  
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;  
Selaginella.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. (bases 1 to 528)

Iturriaga, G. and Cushman, J.C.

An expressed sequence tag (EST) collection from the resurrection  
plant Selaginella lepidophylla

Unpublished

Contact: Cushman JC

Department of Biochemistry

University of Nevada

MS200, Reno, NV 89557-0014, USA

Tel: 775-784-1918

Fax: 775-784-1650

Email: jcushman@unr.edu

PCR PRIMERS

FORWARD: T3 20mer

BACKWARD: T3 21mer

Plate: 005 row: F column: 12

Seq primer: T3 20mer

High quality sequence stop: 528.

Location/Qualifiers

1. 528

## FEATURES

source

/organism="Selaginella lepidophylla"

/mol\_type="mRNA"

/db\_xref="taxon:59777"

/clone="SLA005F12"

/tissue\_type="microphyll fronds undergoing desiccation for  
2.5 h"

/dev\_stage="adult"

/clone\_lib="An expressed sequence tag (EST) collection  
from the resurrection plant Selaginella lepidophylla"

/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site1:  
EcoRI; Site2: XhoI; Library construction was performed  
according to manufacture's (Stratagene, Inc.) recommended  
protocol for the Lambda Uni-zapXR vector and cDNA synthesis  
kit."

BASE COUNT 129 a 125 c 137 g 137 t

## ORIGIN

## Alignment Scores:

Pred. No.: 20.5 Length: 528  
Score: 93.00 Matches: 37  
Percent Similarity: 42.98% Conservative: 15  
Best Local Similarity: 30.58% Mismatches: 43  
Query Match: 9.15% Indels: 26  
DB: 12 Gaps: 4

US-09-965-594-22 (1-197) x BM402566 (1-528)

Qy 94 AspLysAspLeuValGlyTrpGlnAlaProGlnGlySerArgSerLeuThrProCysThr 113

Db 53 GACAAGGATGATAGCGGTGCTGAAGATCGATCTCAAGCACAGATCTCAGCCCATACCC 112

Qy 114 CysGlySerSerAspLeuTyLeuVal----- 122

Db 113 CTTGGAAGTCTGCTCGCATCTGTTGTTGGCCAGAGGTGATGCTATCGGTAAATCTCTTT 172

Qy 123 -----ThrArgHisAlaAspValIleProValArgAlaGlyArgGlyAspSerArg 138

Db 173 GGATTGGATCATACGCTGACACAGCGGTCTATCGTGTCTTCGAAGGAGATTACT--- 229

Qy 139 GlySerLeuLeuSerProArgProIleSerTyLeu----- 150

Db 230 ---TCAGCGGCTAATGGTGGTCCCAATCCAGACGTCATCCAGACAGATGCGGCTATTAA 286

Qy 151 LysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePheArg 170

Db 287 COTGGGAACAGCGGGGTCCGCTATTGACAGATTCTCGAAATTTGATAGGCATCAACACT 346

Qy 171 AlaAlaValSerThrArgGlyValAlaLysAlaValAspPhe---IleProValGluSer 189

Db 347 GCTATATATTTCTCCGCTGGCGCTTCATCAGCGGTGGGCTTTTCCATTCAGTTGACACG 406

Qy 190 Leu 190  
---

```

Db          407 GTT 409
RESULT 15
AQ538021
LOCUS
DEFINITION   560 bp DNA linear GSS 18-MAY-1999
              RPCI-11-32014.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-32014,
              genomic survey sequence.
ACCESSION   AQ538021
VERSION     AQ538021.1 GI:4849711
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
             Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 560)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
             J.C.
TITLE       Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
JOURNAL     Map Building
COMMENT     Unpublished
Other_GSSs: RPCI-11-32014.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.

FEATURES             Location/Qualifiers
     source            1..560
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7622691"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-32014"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      130 a 125 c 181 g 124 t
ORIGIN

Alignment Scores:
Pred. No.:      22.1      Length:      560
Score:          91.00     Matches:      44
Percent Similarity: 37.93%   Conservative: 11
Best Local Similarity: 30.34% Mismatches:   54
Query Match:      9.15%     Indels:      36
DB:              28        Gaps:         6

US-09-965-594-22 (1-197) x AQ538021 (1-560)

QY      36 HisThrGlyArgAspLysAsnGlnValGluGlyGluValGlnIleValSerThrAlaThr 55
      ||| |||||
Db      53 CACCCTGGAAAGGGCTATGAGATCATAGAGGGGCTTCTGCCAGTCTCTCATCCACTGC 112
      ||| |||||
QY      56 GlnThrPheLeuAlaThrSerIleAsnGlyValLeuTrpThrValTyHisGlyAlaGly 75
      ||| |||||
Db      113 CAAGAG-----AATTGCAAAATGTGG-----CACGGATTGTGA 145
      ||| |||||
QY      76 ThrArgThrIleAlaSerProLysGlyProValThrGlnMetTyThrAsp-ValAspLys 95
      ||| |||||
Db      146 GCCAGGATCCAGCTGGCTCTAAAGGG-----AAAACAGTGTCTCACTGTGTCTGCC 196

```

```

QY      95 sAspLeuValGlyTrpGlnAlaProGlnGlySer-----ArgSerLe 109
      ::||| ||||| ||||| |||||
Db      197 AGAGCTTGTGTGCTGGCAACACCCCGAGGAACTTGTGCATGGGGAGGAGTCTGT 256
      ||| |||||
QY      109 uThrProCysThrCysGlySerSerAspLeuTyLeuValThrArgHisAlaAspVal 129
      : |||
Db      257 ACATAGATGT-----CTGGAGCAGAT 277
      |||
QY      129 eProValArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTy 149
      | ||||| ||||| ||||| |||||
Db      278 ATGACGAGCCATGGAGGAGAGAGAGAGCCCTTCTCCCTGGCCCGCCCTGGACCATA 337
      ||||| ||||| ||||| |||||
QY      149 rLeuLysGlySer-SerGlyGlyProLeuLeuCysProLaGlyHisAlaValGlyIleP 169
      ||||| ||||| ||||| |||||
Db      338 TCTGAAAGCTGTGAAGTGGAGGC-----TGC CGCGGGGGAGCTGATGCAGCTGTTTC 388
      ||||| ||||| ||||| |||||
QY      169 heArgAlaAla 172
      |||
Db      389 TCACAGGGGCT 399

```

Search completed: August 31, 2003, 04:27:50  
Job time : 1914.31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 17:42:58 ; Search time 2.49162 Seconds  
(without alignments)  
700.745 Million cell updates/sec

Title: US-09-965-594-26  
Perfect score: 49  
Sequence: 1 GSWIVGRIVL 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	11	21	AA15227 Hepatitis C virus
2	49	100.0	12	23	AA48242 Hepatitis C virus
3	49	100.0	13	22	AA96864 Hepatitis C virus
4	49	100.0	23	19	AAW40552 Synthetic nonstruc
5	49	100.0	23	22	AAW52536 Peptide KKS4a use
6	49	100.0	23	22	AA664158 KKS4a peptide. S
7	49	100.0	23	22	AA67391 KKS4a peptide. S
8	49	100.0	23	22	AA66371 Hepatitis C virus
9	49	100.0	23	23	ABG32508 Peptide KKS4a for

10	49	100.0	23	23	ABG31914 KKS4a peptide. S
11	49	100.0	23	23	AAE18687 NS4A peptide used
12	49	100.0	23	23	AAU76376 Hepatitis C virus
13	49	100.0	23	24	ABG72264 Hepatitis C virus
14	49	100.0	195	21	AA15212 Hepatitis C virus
15	49	100.0	195	21	AA15220 Hepatitis C virus
16	49	100.0	200	13	AA23846 HCV NS2-NS4 peptid
17	48	98.0	12	21	AA144731 Hepatitis C virus
18	48	98.0	13	22	AA96862 Hepatitis C virus
19	48	98.0	14	18	AAW13792 Hepatitis C virus
20	48	98.0	14	22	AAW74387 NS3 protease activ
21	48	98.0	14	22	AA92337 Virus related pept
22	48	98.0	16	21	AA54448 Peptide 4A4 used t
23	48	98.0	16	22	AA96851 Hepatitis C virus
24	48	98.0	16	22	AA96853 Hepatitis C virus
25	48	98.0	16	22	AA96854 Hepatitis C virus
26	48	98.0	17	21	AA99552 Hepatitis C virus
27	48	98.0	17	21	AA93773 HCV NS3A cofactor
28	48	98.0	17	21	AA93775 HCV NS3A cofactor
29	48	98.0	17	22	AA97114 Hepatitis C virus
30	48	98.0	18	23	ABW05367 NS4a peptide. Hep
31	48	98.0	23	20	AA15763 Substrate peptide
32	48	98.0	23	21	AA23810 Synthetic peptide
33	48	98.0	23	22	AA96855 Hepatitis C virus
34	48	98.0	23	22	AA92336 Virus related pept
35	48	98.0	28	19	AAW37386 Hepatitis C virus
36	48	98.0	32	22	AA96856 Hepatitis C virus
37	48	98.0	34	16	AAW2856 NS3 serine proteas
38	48	98.0	36	19	AAW50782 Peptide used in im
39	48	98.0	54	16	AAW2855 NS4A protein. Hep
40	48	98.0	54	19	AAW37808 Nonstructural doma
41	48	98.0	54	20	AAW17898 Native HCV NS4A pe
42	48	98.0	63	15	AAW49651 HCV peptide C14-1.
43	48	98.0	86	18	AAW09051 Hepatitis C virus
44	48	98.0	87	15	AAW49652 HCV peptide C14-1.
45	48	98.0	87	17	AAW95545 HCV II chimeric ep

ALIGNMENTS

RESULT 1  
AA15227  
ID AA15227 standard; protein; 11 AA.  
XX  
AC AA15227;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE Hepatitis C virus NS4A residues 21-31.  
XX  
XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer.  
XX  
OS Hepatitis C virus.  
XX  
XX W0200040707-A1.  
XX  
PD 13-JUL-2000.  
XX  
PF 06-JAN-2000; 2000WO-US00345.  
XX  
PR 08-JAN-1999; 99US-0115271.  
XX  
FA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
XX  
DR WPI; 2000-465976/40.  
XX  
PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C



```

AC AAW40552;
XX
XX
XX 20-NOV-1998 (first entry)
XX
XX Synthetic nonstructural peptide SNS4A.
XX
XX Synthetic Hepatitis C nonstructural protein; SNS4A peptide;
XX cofactor; NS3 protease.
XX
XX OS Synthetic.
XX OS Hepatitis C virus.
XX
XX PN WO9811134-A1.
XX
XX PD 19-MAR-1998.
XX
XX PF 12-SEP-1997; 97WO-US16182.
XX
XX PR 18-OCT-1996; 96US-0731336.
XX PR 12-SEP-1996; 96US-0025274.
XX
XX PA (VERT-) VERTEX PHARM INC.
XX
XX PI Fox T, Kim JL, Lin C, Morgenstern KA, Thomson JA;
XX
XX DR WPI; 1998-250953/22.
XX
XX
XX New hepatitis C virus crystal compositions - comprising a HCV
XX PT NS3-like polypeptide complexed with a NS4A-like polypeptide, used
XX PT particularly for drug design
XX
XX PS Claim 4; Page 30; 97pp; English.
XX
XX This is the amino acid sequence of the novel SNS4A (synthetic
XX CC Hepatitis C nonstructural protein 4A) peptide. It acts as a cofactor
XX CC for the NS3 protease in order to achieve proteolytic processing of
XX CC Hepatitis C virus (HCV) nonstructural proteins. It is used in the
XX CC method of the invention as part of a device which can be used to
XX CC provide information for the design of drugs for the treatment of HCV
XX CC infection. They can also be used for determining the 3-dimensional
XX CC structure of molecules or molecular complexes which contain at least
XX CC some structurally similar features to a HCV NS3 serine protease domain.
XX
XX SQ Sequence 23 AA;
XX
XX Query Match 100.0%; Score 49; DB 19; Length 23;
XX Best Local Similarity 100.0%; Pred. No. 0.055;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11
Db 3 GSVVIVGRIVL 13

RESULT 5
AAM52536
ID AAM52536 standard; peptide; 23 AA.
XX
XX AC AAM52536;
XX
XX 31-JAN-2002 (first entry)
XX
XX DE Peptide KNS4a used in an enzyme assay.
XX
XX KW Virucide; pyrrolopyrazinone derivative; Hepatitis C virus inhibitor;
XX KW nonstructural 3 protease; NS3 protease; viral infection.
XX
XX OS Synthetic.
XX
XX PN WO200164678-A2.
XX
XX PD 07-SEP-2001.
XX

AAW40552;
XX
XX 28-FEB-2001; 2001WO-US06269.
XX
XX PR 29-FEB-2000; 2000US-185618P.
XX
XX PA (DUPO ) DUPONT PHARM CO.
XX
XX Zhang X, Han W;
XX
XX WPI; 2001-656752/75.
XX
XX New pyrrolopyrazinone derivatives useful for treating Hepatitis C virus
XX infection are NS3 protease inhibitors -
XX
XX PS Disclosure; Page 130; 191pp; English.
XX
XX The present invention relates to a novel pyrrolopyrazinone derivative,
XX its stereoisomer or salt. It was found that the derivative is a Hepatitis
XX C virus (HCV) nonstructural (NS) 3 protease inhibitor. The derivative is
XX useful for the manufacture of a medicament for the treatment of HCV and
XX in therapy for treating HCV infection. The present peptide was used to
XX illustrate the present invention.
XX
XX SQ Sequence 23 AA;
XX
XX Query Match 100.0%; Score 49; DB 22; Length 23;
XX Best Local Similarity 100.0%; Pred. No. 0.055;
XX Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11
Db 3 GSVVIVGRIVL 13

RESULT 6
AAG64158
ID AAG64158 standard; peptide; 23 AA.
XX
XX AC AAG64158;
XX
XX DT 19-OCT-2001 (first entry)
XX
XX DE KNS4a peptide.
XX
XX KW Hepatitis C virus; HCV; NS3 protease; alpha-ketoamide inhibitor;
XX KW virucide; hepatotropic; antiinflammatory; viral infection; KNS4a.
XX
XX OS Synthetic.
XX
XX PN WO200140262-A1.
XX
XX PD 07-JUN-2001.
XX
XX PF 01-DEC-2000; 2000WO-US32677.
XX
XX PR 03-DEC-1999; 99US-0168998.
XX
XX PA (DUPO ) DU PONT PHARM CO.
XX
XX PI Han W;
XX
XX WPI; 2001-464936/50.
XX
XX New ketoamide derivatives useful for treating infections e.g. hepatitis
XX C virus -
XX
XX PS Disclosure; Page 195; 282pp; English.
XX
XX The invention relates to novel ketoamide and ketoester derivatives
XX for use as inhibitors of hepatitis C virus (HCV) NS3 protease inhibitors.
XX The compounds are useful for treating viral infections e.g. hepatitis C
XX virus. The present sequence was used in an experiment measuring the
XX effect of an inhibitor on the rate of hydrolysis of an ester substrate.
XX

```

SQ Sequence 23 AA:  
 Query Match 100.0%; Score 49; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.055;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11  
 |||||  
 DB 3 GSVWIVGRIVL 13

RESULT 7  
 AAB67391  
 ID AAB67391 standard; peptide: 23 AA.  
 AC AAB67391;  
 DT 26-APR-2001 (first entry)  
 XX  
 DE KKN54a peptide.  
 KW Lactam; hepatitis C virus; HCV; NS3 protease.  
 XX Synthetic.  
 OS  
 PN WO200107407-A1.  
 XX  
 PD 01-FEB-2001.  
 XX  
 PF 26-JUL-2000; 2000WO-US20189.  
 XX  
 PR 26-JUL-1999; 99US-0145631.  
 XX  
 PA (DUPO ) DU PONT PHARM CO.  
 XX  
 PI Priestley ES, Decicco CP;  
 XX  
 DR WPI; 2001-159696/16.  
 XX  
 PT New lactam derivatives are hepatitis C virus NS3 protease inhibitors  
 useful for treating HCV infections .  
 XX  
 PS Example 26; Page 100; 130pp; English.  
 XX  
 CC The present invention relates to Lactam derivatives. These derivatives  
 may be used for treating hepatitis C virus (HCV) infection. They can  
 also be used for inhibiting HCV in a body fluid sample and as a  
 standard or reagent in a test or assay for determining the ability  
 of a potential pharmaceutical to inhibit HCV NS3 protease and/or HCV  
 growth.  
 XX  
 SQ Sequence 23 AA;  
 Query Match 100.0%; Score 49; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.055;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11  
 |||||  
 DB 3 GSVWIVGRIVL 13

RESULT 8  
 AAB66371  
 ID AAB66371 standard; peptide: 23 AA.  
 AC AAB66371;  
 DT 09-APR-2001 (first entry)  
 XX  
 DE Hepatitis C virus protease inhibitor related peptide #1.  
 XX  
 KW Hepatitis C virus; protease; boronic acid; inhibitor; liver cirrhosis;

KW liver cancer; NS3; antiviral agent.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200102424-A2.  
 XX  
 PD 11-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000WO-US18655.  
 XX  
 PR 07-JUL-1999; 99US-0142561.  
 XX  
 PA (DUPO ) DU PONT PHARM CO.  
 XX  
 PI Kettner CA, Jagannathan S, Forsyth TP;  
 XX  
 DR WPI; 2001-103001/11.  
 XX  
 PT New boronic acid derivatives, optionally containing peptides, used to  
 treat hepatitis C infections, are hepatitis C viral protease inhibitors  
 .  
 XX  
 PS Example 60; Page 208; 258pp; English.  
 XX  
 CC The present invention provides a number of boronic acid derivatives which  
 act as inhibitors of the hepatitis C virus NS3 protease enzyme. They can  
 be used to treat infection by the virus, which can cause liver cirrhosis  
 and liver cancer.  
 XX  
 SQ Sequence 23 AA;  
 Query Match 100.0%; Score 49; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.055;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVWIVGRIVL 11  
 |||||  
 DB 3 GSVWIVGRIVL 13

RESULT 9  
 ABG32508  
 ID ABG32508 standard; peptide: 23 AA.  
 XX  
 AC ABG32508;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Peptide KKN54a for HCV NS3 protease kinetic assay.  
 XX  
 KW NS3; HCV; protease; HCV infection; hepatitis; cirrhosis; liver cancer;  
 KW pyrimidinone; serine protease inhibitor; virucide; hepatotropic;  
 KW antiinflammatory; blood plasma processing; KKN54a.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200248116-A2.  
 XX  
 PD 20-JUN-2002.  
 XX  
 PF 12-DEC-2001; 2001WO-US47911.  
 XX  
 PR 13-DEC-2000; 2000US-255290P.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB PHARMA CO.  
 XX  
 PI Glunz PW, Douty BD, Han W;  
 XX  
 DR WPI; 2002-627251/67.  
 XX  
 PT New pyrimidinones useful as serine protease inhibitors in the treatment  
 of e.g. viral infection .  
 XX

Example 140; Page 192; 270pp; English.

The invention relates to pyrimidinones of a formula given in the claims of the specification, their stereoisomers, salts and prodrugs. In assays, the pyrimidinone compounds inhibited Hepatitis C virus (HCV) NS3 protease with IC<sub>50</sub> values of less than 100 micro M. The compounds are useful for treating viral infection e.g. HCV infection (the causative agent of acute hepatitis and associated with cirrhosis and liver cancer) and as a reagent used as inhibitors of HCV protease in the processing of blood plasma for diagnostic and other commercial purposes. The present sequence is a peptide, KNNS4a, used in an NS3 kinetic assay.

Sequence 23 AA:

Query Match 100.0%; Score 49; DB 23; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11  
          |||||  
DB 3 GSVVIVGRIVL 13

RESULT 10

ABG31914  
ID ABG31914 standard; Peptide; 23 AA.

XX AC ABG31914;

XX DT 05-NOV-2002 (first entry)

XX DE KNNS4a peptide.

XX HCV; hepatitis C; imidazolidinone; serine protease inhibitor;  
XX virucide; hepatotropic; antiinflammatory; NS3 protease; KNNS4a;  
XX growth inhibitor; viral infection; blood plasma processing.

XX OS Synthetic.

XX PN WO200248157-A2.

XX PD 20-JUN-2002.

XX PF 12-DEC-2001; 2001WO-US47916.

XX PR 13-DEC-2000; 2000US-255168P.

XX PA (BRIM ) BRISTOL MYERS SQUIBB PHARMA CO.

XX PI Han Q;

XX DR WPI; 2002-599498/64.

XX PT New imidazolidinones useful as serine protease inhibitors in the treatment of e.g. viral infection

XX PS Example 20; Page 112; 173pp; English.

XX CC This invention relates to novel imidazolidinones or their stereoisomers, salts or prodrugs which are useful as serine protease inhibitors. The imidazolidinones of the invention may have virucide, hepatotropic, or antiinflammatory activities and may be used as a serine protease inhibitor (preferably Hepatitis C virus (HCV) NS3 protease inhibitor) or a HCV growth inhibitor. Compounds of the invention are useful for treating viral infection e.g. hepatitis C virus (HCV) infection and as a reagent used as inhibitors of HCV protease in the processing of blood plasma for diagnostic and other commercial purposes. The imidazolidinones of the invention inhibit HCV NS3 protease and/or HCV growth and thus can be used in the blood plasma assay. The present sequence represents the KNNS4a peptide used in enzyme assay experiments in the examples of the specification.

XX Sequence 23 AA:

Query Match 100.0%; Score 49; DB 23; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11  
          |||||  
DB 3 GSVVIVGRIVL 13

RESULT 11

AAE18687  
ID AAE18687 standard; peptide; 23 AA.

XX AC AAE18687;

XX DT 17-MAY-2002 (first entry)

XX DE NS4A peptide used to purify NS3/4a conformational epitope.

XX KW Hepatitis C virus; NS3/4a antibody; HCV infection; NS4A peptide.

XX OS Unidentified.

XX PN WO200196875-A2.

XX PD 20-DEC-2001.

XX PF 14-JUN-2001; 2001WO-US19369.

XX PR 15-JUN-2000; 2000US-212082P.

XX PR 02-APR-2001; 2001US-280811P.

XX PR 02-APR-2001; 2001US-280867P.

XX PA (CHIR ) CHIRON CORP.

XX PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

XX PI Medina-Selby A;

XX DR WPI; 2002-179522/23.

XX PT Immunocassay solid support useful for detecting hepatitis C virus infection in a biological sample, comprises at least one of HCV

XX PI anti-core antibody and HCV NS3/4a epitope, bound to the support

XX PS Example 2; Page 50; 87pp; English.

XX CC The present invention relates to hepatitis C virus (HCV) core antigen and NS (nonstructural) 3/4a antibody combination assay that can detect both HCV antigens and antibodies present in a sample using a single solid matrix as well as immunocassay solid supports for use in the assay. The solid support is useful for detecting HCV infection in a biological sample. The present sequence is NS4A peptide which is used to purify NS3/4a conformational epitope in the exemplification of the invention.

XX SQ Sequence 23 AA:

Query Match 100.0%; Score 49; DB 23; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSVVIVGRIVL 11  
          |||||  
DB 3 GSVVIVGRIVL 13

RESULT 12

AAU76376  
ID AAU76376 standard; Peptide; 23 AA.

XX AC AAU76376;

XX DT 08-MAY-2002 (first entry)



```

XX Hepatitis C virus (non-structural protein) NS4A peptide sequence.
DE
XX Hepatitis C virus: HCV; NS3/4a conformational epitope; seroconversion;
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein; NS4A.
XX
OS Hepatitis C virus.
XX
PN WO200196870-A2.
PD
XX 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-US19156.
XX
XX 15-JUN-2000; 2000US-212082P.
PR
XX 02-APR-2001; 2001US-280811P.
PR
XX 02-APR-2001; 2001US-280867P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
PI Medina-selby A;
PI
XX WPI; 2002-090228/12.
XX
XX Immunoassay solid support, useful for detecting hepatitis C virus
PT infection in biological sample, comprises HCV NS3/4a conformational
PT epitope and multiple epitope fusion antigen bound to the support -
XX
XX Example 3; Page 48; 92pp; English.
XX
XX The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the non-structural protein NS4A peptide sequence. The peptide
CC was used in the invention for the purification of NS3/4a conformational
CC epitope.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 49; DB 23; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GSVIVIGRVL 11
Dd | | | | | | | | | |
3 GSVIVIGRVL 13
RESULT 13
ABG72264
ID ABG72264 standard; peptide; 23 AA.
XX
XX ABG72264;
AC
XX DT- 06-MAR-2003 (first entry)
XX

```

```

XX Hepatitis C Virus type-1 (HCV-1) NS4a peptide.
DE
XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1;
KW NS3/4a conformational epitope; multiple epitope fusion antigen;
KW MEFA; anti-HCV antibody; NS3/4a conformational antigen;
KW HCV infection; E2 hypervariable region.
XX
XX Hepatitis C virus type 1.
OS
XX
XX US2002146685-A1.
PN
XX 10-OCT-2002.
PD
XX
XX 14-JUN-2001; 2001US-0881654.
XX
XX 15-JUN-2000; 2000US-212082P.
PR
XX 02-APR-2001; 2001US-280811P.
PR
XX 02-APR-2001; 2001US-280867P.
XX
XX (CHIE/) CHIEN D Y.
XX
XX (ARCA/) ARCANGEL P.
XX
XX (TAND/) TANDESKE L.
XX
XX (GEOR/) GEORGE-NASCIMENTO C.
XX
XX (COIT/) COIT D.
XX
XX (MEDI/) MEDINA-SELBY A.
XX
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Selby A;
PI
XX WPI; 2003-147573/14.
XX
XX Immunoassay solid support for detecting Hepatitis C Virus infection in
PT biological samples, comprises Hepatitis C Virus conformational epitope
PT and multiple epitope fusion antigen -
XX
XX Example 3; Page 17; 45pp; English.
XX
XX The present invention relates to immunoassays comprising Hepatitis C
CC virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC the multiple epitope fusion antigen react with anti-HCV antibodies
CC present in a biological sample from an HCV-infected individual. The
CC immunoassays and methods of the invention are useful for detecting
CC HCV infection in a biological sample. The inventive immunoassay solid
CC support provides a sensitive and reliable method for detecting early
CC HCV seroconversion. The assays can detect HCV infection caused by any
CC six known genotypes of HCV. The use of the multiple epitope fusion
CC proteins decreases masking problems, improves sensitivity in detecting
CC antibodies by allowing a greater number of epitopes on a unit area
CC of substrate, and improves selectivity. The present sequence
CC representing HCV type 1 (HCV-1) NS4a peptide is used in a protease
CC enzyme activity assay in the examples of the present invention.
XX
XX Sequence 23 AA;
SQ
Query Match 100.0%; Score 49; DB 24; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GSVIVIGRVL 11
Dd | | | | | | | | | |
3 GSVIVIGRVL 13
RESULT 14
AAB15212
ID AAB15212 standard; protein; 195 AA.
XX
XX AAB15212;
AC
XX DT 19-DEC-2000 (first entry)
XX

```

DE Hepatitis C virus NS4A-NS3 fusion protease #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 DR N-PSDB; AAA73328.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT .  
 XX  
 PS Example 2; Fig 10; 66pp; English.  
 XX  
 CC The present sequence is a fusion protein created using the Hepatitis C  
 CC virus (HCV) NS3 and NS4A protease enzymes. These proteins are both  
 CC essential for the replication of the virus, acting to cleave its  
 CC replicative proteins from the polyprotein produced from the HCV genome.  
 CC Inhibitors of the two proteins should be effective as antiviral  
 CC treatments of HCV infection. This is useful as HCV can lead to chronic  
 CC liver disease such as cirrhosis, liver failure and liver cancer. The  
 CC present invention concerns a number of NS3 mutants and NS3-NS4A fusion  
 CC proteins which can be used to identify inhibitors of this type, as well  
 CC as enabling structural studies of the protease and protease:inhibitor  
 CC complexes.  
 XX  
 SQ Sequence 195 AA;  
 Query Match 100.0%; Score 49; DB 21; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 0.56;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSVVIVGRIVL 11  
 DB 5 GSVVIVGRIVL 15  
 RESULT 15  
 AAB15220  
 ID AAB15220 standard; protein; 195 AA.  
 XX  
 AC AAB15220;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease #2.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 PN WO200040707-A1.  
 XX  
 PD 13-JUL-2000.  
 XX

PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 DR N-PSDB; AAA73329.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT .  
 XX  
 PS Claim 23; Fig 12; 66pp; English.  
 XX  
 CC The present sequence is a mutated version of a fusion protein created  
 CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
 CC proteins are both essential for the replication of the virus, acting to  
 CC cleave its replicative proteins from the polyprotein produced from the  
 CC HCV genome. Inhibitors of the two proteins should be effective as  
 CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
 CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
 CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
 CC fusion proteins which can be used to identify inhibitors of this type, as  
 CC well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes. This sequence contains the alpha-helix0-1  
 CC variant.  
 XX  
 SQ Sequence 195 AA;  
 Query Match 100.0%; Score 49; DB 21; Length 195;  
 Best Local Similarity 100.0%; Pred. No. 0.56;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSVVIVGRIVL 11  
 DB 5 GSVVIVGRIVL 15  
 Search completed: August 30, 2003, 19:12:26  
 Job time : 3.49162 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 30, 2003, 19:02:22 ; Search time 0.905317 Seconds  
(without alignments)  
1168.492 Million cell updates/sec

Title: US-09-965-594-26

Perfect score: 49

Sequence: 1 GSVVIVGRIVL 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	3010	1 GNMVTC	genome polyprotein
2	48	98.0	3010	1 GNMVTC	genome polyprotein
3	48	98.0	3010	1 A45573	genome polyprotein
4	48	98.0	3010	1 S18030	genome polyprotein
5	48	98.0	3010	1 GNMVTC	genome polyprotein
6	44	89.8	492	2 PS0126	polyprotein - hepa
7	44	89.8	876	2 PC2219	polypeptide - hepa
8	44	89.8	3011	1 GNMVTC	genome polyprotein
9	44	89.8	3011	1 S40770	genome polyprotein
10	44	89.8	3014	1 JC5620	genome polyprotein
11	43	87.8	3011	1 GNMVTC	genome polyprotein
12	42	85.7	716	2 J01366	polyprotein - hepa
13	37	75.5	718	2 G70978	probable copper-tr
14	35	71.4	574	2 T41395	probable dna polym
15	35	71.4	876	2 E89949	valine-tRNA ligase
16	34	69.4	44	2 PS0117	H-2 class I histoc
17	34	69.4	196	2 S54580	probable membrane
18	34	69.4	224	2 C46357	env polyprotein -
19	34	69.4	570	2 E71234	hypothetical prote
20	34	69.4	649	2 T05630	hypothetical prote
21	34	69.4	859	1 VCLJEW	env polyprotein pr
22	34	69.4	859	1 VCLJ22	env polyprotein pr
23	34	69.4	859	1 VCLJEW	env polyprotein pr
24	34	69.4	859	1 VCLJEW	env polyprotein pr
25	34	69.4	859	1 VCLJ22	env polyprotein pr
26	34	69.4	859	1 VCLJ22	env polyprotein pr
27	34	69.4	859	1 VCLJ22	env polyprotein pr
28	34	69.4	860	1 VCLJ24	env polyprotein pr
29	33	67.3	37	2 PS0130	H-2 class I histoc

30	33	67.3	37	2 PS0127	H-2 class I histoc
31	33	67.3	38	2 PS0118	H-2 class I histoc
32	33	67.3	39	2 A32934	H-2 class I-like h
33	33	67.3	142	2 PC1307	genome polyprotein
34	33	67.3	209	2 PC1306	genome polyprotein
35	33	67.3	220	2 D75611	conserved hypothe
36	33	67.3	226	2 D90908	probable tail asse
37	33	67.3	248	2 AG3213	3-oxoacyl-(acyl-ca
38	33	67.3	271	2 F95091	conserved domain p
39	33	67.3	291	2 B97959	conserved hypothe
40	33	67.3	300	2 C85631	hypothetical prote
41	33	67.3	316	2 F69491	methyltetrahydro
42	33	67.3	331	2 F89771	lipoprotein (impor
43	33	67.3	386	2 B71407	hypothetical prote
44	33	67.3	395	2 E90047	hypothetical prote
45	33	67.3	405	2 C83204	argininosuccinate

## ALIGNMENTS

## RESULT 1

## GNMVT

genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J)  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001  
C:Accession: A38465  
R:Rakamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from hu  
A:Reference number: A38465; MUID:91140698; PMID:1847440  
A:Accession: A38465

A:Molecule type: genomic RNA

A:Residues: 1-3010 <TAK>

A:Cross-references: EMBL:M58335; NID:g329770; PID:AAA72945.1; PID:g329771

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MES>

F:330-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus (strain J) #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,20:

Query Match 98.0%; Score 48; DB 1; Length 3010;

Best Local Similarity 90.9%; Pred. No. 2.4;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11

|||||||

Db 1678 GSVVIVGRIVL 1688

## RESULT 2

## GNMVT

genome polyprotein - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstruc  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001  
C:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; S

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patie

```
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65A, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variations
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KAT2>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2401,2402,2403,2404,2405,2406,2407,2408,2409,2410,2411,2412,2413,2414,2415,2416,2417,2418,2419,2420,2421,2422,2423,2424,2425,2426,2427,2428,2429,2430,2431,2432,2433,2434,2435,2436,2437,2438,2439,2440,2441,2442,2443,2444,2445,2446,2447,2448,2449,2450,2451,2452,2453,2454,2455,2456,2457,2458,2459,2460,2461,2462,2463,2464,2465,2466,2467,2468,2469,2470,2471,2472,2473,2474,2475,2476,2477,2478,2479,2480,2481,2482,2483,2484,2485,2486,2487,2488,2489,2490,2491,2492,2493,2494,2495,2496,2497,2498,2499,2500,2501,2502,2503,2504,2505,2506,2507,2508,2509,2510,2511,2512,2513,2514,2515,2516,2517,2518,2519,2520,2521,2522,2523,2524,2525,2526,2527,2528,2529,2530,2531,2532,2533,2534,2535,2536,2537,2538,2539,2540,2541,2542,2543,2544,2545,2546,2547,2548,2549,2550,2551,2552,2553,2554,2555,2556,2557,2558,2559,2560,2561,2562,2563,2564,2565,2566,2567,2568,2569,2570,2571,2572,2573,2574,2575,2576,2577,2578,2579,2580,2581,2582,2583,2584,2585,2586,2587,2588,2589,2590,2591,2592,2593,2594,2595,2596,2597,2598,2599,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,2610,2611,2612,2613,2614,2615,2616,2617,2618,2619,2620,2621,2622,2623,2624,2625,2626,2627,2628,2629,2630,2631,2632,2633,2634,2635,2636,2637,2638,2639,2640,2641,2642,2643,2644,2645,2646,2647,2648,2649,2650,2651,2652,2653,2654,2655,2656,2657,2658,2659,2660,2661,2662,2663,2664,2665,2666,2667,2668,2669,2670,2671,2672,2673,2674,2675,2676,2677,2678,2679,2680,2681,2682,2683,2684,2685,2686,2687,2688,2689,2690,2691,2692,2693,2694,2695,2696,2697,2698,2699,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,2710,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2741,2742,2743,2744,2745,2746,2747,2748,2749,2750,2751,2752,2753,2754,2755,2756,2757,2758,2759,2760,2761,2762,2763,2764,2765,2766,2767,2768,2769,2770,2771,2772,2773,2774,2775,2776,2777,2778,2779,2780,2781,2782,2783,2784,2785,2786,2787,2788,2789,2790,2791,2792,2793,2794,2795,2796,2797,2798,2799,2800,2801,2802,2803,2804,2805,2806,2807,2808,2809,2810,2811,2812,2813,2814,2815,2816,2817,2818,2819,2820,2821,2822,2823,2824,2825,2826,2827,2828,2829,2830,2831,2832,2833,2834,2835,2836,2837,2838,2839,2840,2841,2842,2843,2844,2845,2846,2847,2848,2849,2850,2851,2852,2853,2854,2855,2856,2857,2858,2859,2860,2861,2862,2863,2864,2865,2866,2867,2868,2869,2870,2871,2872,2873,2874,2875,2876,2877,2878,2879,2880,2881,2882,2883,2884,2885,2886,2887,2888,2889,2890,2891,2892,2893,2894,2895,2896,2897,2898,2899,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,2924,2925,2926,2927,2928,2929,2930,2931,2932,2933,2934,2935,2936,2937,2938,2939,2940,2941,2942,2943,2944,2945,2946,2947,2948,2949,2950,2951,2952,2953,2954,2955,2956,2957,2958,2959,2960,2961,2962,2963,2964,2965,2966,2967,2968,2969,2970,2971,2972,2973,2974,2975,2976,2977,2978,2979,2980,2981,2982,2983,2984,2985,2986,2987,2988,2989,2990,2991,2992,2993,2994,2995,2996,2997,2998,2999,3000,3001,3002,3003,3004,3005,3006,3007,3008,3009,3010,3011,3012,3013,3014,3015,3016,3017,3018,3019,3020,3021,3022,3023,3024,3025,3026,3027,3028,3029,3030,3031,3032,3033,3034,3035,3036,3037,3038,3039,3040,3041,3042,3043,3044,3045,3046,3047,3048,3049,3050,3051,3052,3053,3054,3055,3056,3057,3058,3059,3060,3061,3062,3063,3064,3065,3066,3067,3068,3069,3070,3071,3072,3073,3074,3075,3076,3077,3078,3079,3080,3081,3082,3083,3084,3085,3086,3087,3088,3089,3090,3091,3092,3093,3094,3095,3096,3097,3098,3099,3100,3101,3102,3103,3104,3105,3106,3107,3108,3109,3110,3111,3112,3113,3114,3115,3116,3117,3118,3119,3120,3121,3122,3123,3124,3125,3126,3127,3128,3129,3130,3131,3132,3133,3134,3135,3136,3137,3138,3139,3140,3141,3142,3143,3144,3145,3146,3147,3148,3149,3150,3151,3152,3153,3154,3155,3156,3157,3158,3159,3160,3161,3162,3163,3164,3165,3166,3167,3168,3169,3170,3171,3172,3173,3174,3175,3176,3177,3178,3179,3180,3181,3182,3183,3184,3185,3186,3187,3188,3189,3190,3191,3192,3193,3194,3195,3196,3197,3198,3199,3200,3201,3202,3203,3204,3205,3206,3207,3208,3209,3210,3211,3212,3213,3214,3215,3216,3217,3218,3219,3220,3221,3222,3223,3224,3225,3226,3227,3228,3229,3230,3231,3232,3233,3234,3235,3236,3237,3238,3239,3240,3241,3242,3243,3244,3245,3246,3247,3248,3249,3250,3251,3252,3253,3254,3255,3256,3257,3258,3259,3260,3261,3262,3263,3264,3265,3266,3267,3268,3269,3270,3271,3272,3273,3274,3275,3276,3277,3278,3279,3280,3281,3282,3283,3284,3285,3286,3287,3288,3289,3290,3291,3292,3293,3294,3295,3296,3297,3298,3299,3300,3301,3302,3303,3304,3305,3306,3307,3308,3309,3310,3311,3312,3313,3314,3315,3316,3317,3318,3319,3320,3321,3322,3323,3324,3325,3326,3327,3328,3329,3330,3331,3332,3333,3334,3335,3336,3337,3338,3339,3340,3341,3342,3343,3344,3345,3346,3347,3348,3349,3350,3351,3352,3353,3354,3355,3356,3357,3358,3359,3360,3361,3362,3363,3364,3365,3366,3367,3368,3369,3370,3371,3372,3373,3374,3375,3376,3377,3378,3379,3380,3381,3382,3383,3384,3385,3386,3387,3388,3389,3390,3391,3392,3393,3394,3395,3396,3397,3398,3399,3400,3401,3402,3403,3404,3405,3406,3407,3408,3409,3410,3411,3412,3413,3414,3415,3416,3417,3418,3419,3420,3421,3422,3423,3424,3425,3426,3427,3428,3429,3430,3431,3432,3433,3434,3435,3436,3437,3438,3439,3440,3441,3442,3443,3444,3445,3446,3447,3448,3449,3450,3451,3452,3453,3454,3455,3456,3457,3458,3459,3460,3461,3462,3463,3464,3465,3466,3467,3468,3469,3470,3471,3472,3473,3474,3475,3476,3477,3478,3479,3480,3481,3482,3483,3484,3485,3486,3487,3488,3489,3490,3491,3492,3493,3494,3495,3496,3497,3498,3499,3500,3501,3502,3503,3504,3505,3506,3507,3508,3509,3510,3511,3512,3513,3514,3515,3516,3517,3518,3519,3520,3521,3522,3523,3524,3525,3526,3527,3528,3529,3530,3531,3532,3533,3534,3535,3536,3537,3538,3539,3540,3541,3542,3543,3544,3545,3546,3547,3548,3549,3550,3551,3552,3553,3554,3555,3556,3557,3558,3559,3560,3561,3562,3563,3564,3565,3566,3567,3568,3569,3570,3571,3572,3573,3574,3575,3576,3577,3578,3579,3580,3581,3582,3583,3584,3585,3586,3587,3588,3589,3590,3591,3592,3593,3594,3595,3596,3597,3598,3599,3600,3601,3602,3603,3604,3605,3606,3607,3608,3609,3610,3611,3612,3613,3614,3615,3616,3617,3618,3619,3620,3621,3622,3623,3624,3625,3626,3627,3628,3629,3630,3631,3632,3633,3634,3635,3636,3637,3638,3639,3640,3641,3642,3643,3644,3645,3646,3647,3648,3649,3650,3651,3652,3653,3654,3655,3656,3657,3658,3659,3660,3661,3662,3663,3664,3665,3666,3667,3668,3669,3670,3671,3672,3673,3674,3675,3676,3677,3678,3679,3680,3681,3682,3683,3684,3685,3686,3687,3688,3689,3690,3691,3692,3693,3694,3695,3696,3697,3698,3699,3700,3701,3702,3703,3704,3705,3706,3707,3708,3709,3710,3711,3712,3713,3714,3715,3716,3717,3718,3719,3720,3721,3722,3723,3724,3725,3726,3727,3728,3729,3730,3731,3732,3733,3734,3735,3736,3737,3738,3739,3740,3741,3742,3743,3744,3745,3746,3747,3748,3749,3750,3751,3752,3753,3754,3755,3756,3757,3758,3759,3760,3761,3762,3763,3764,3765,3766,3767,3768,3769,3770,3771,3772,3773,3774,3775,3776,3777,3778,3779,3780,3781,3782,3783,3784,3785,3786,3787,3788,3789,3790,3791,3792,3793,3794,3795,3796,3797,3798,3799,3800,3801,3802,3803,3804,3805,3806,3807,3808,3809,3810,3811,3812,3813,3814,3815,3816,3817,3818,3819,3820,3821,3822,3823,3824,3825,3826,3827,3828,3829,3830,3831,3832,3833,3834,3835,3836,3837,3838,3839,3840,3841,3842,3843,3844,3845,3846,3847,3848,3849,3850,3851,3852,3853,3854,3855,3856,3857,3858,3859,3860,3861,3862,3863,3864,3865,3866,3867,3868,3869,3870,3871,3872,3873,3874,3875,3876,3877,3878,3879,3880,3881,3882,3883,3884,3885,3886,3887,3888,3889,3890,3891,3892,3893,3894,3895,3896,3897,3898,3899,3900,3901,3902,3903,3904,3905,3906,3907,3908,3909,3910,3911,3912,3913,3914,3915,3916,3917,3918,3919,3920,3921,3922,3923,3924,3925,3926,3927,3928,3929,3930,3931,3932,3933,3934,3935,3936,3937,3938,3939,3940,3941,3942,3943,3944,3945,3946,3947,3948,3949,3950,3951,3952,3953,3954,3955,3956,3957,3958,3959,3960,3961,3962,3963,3964,3965,3966,3967,3968,3969,3970,3971,3972,3973,3974,3975,3976,3977,3978,3979,3980,3981,3982,3983,3984,3985,3986,3987,3988,3989,3990,3991,3992,3993,3994,3995,3996,3997,3998,3999,4000,4001,4002,4003,4004,4005,4006,4007,4008,4009,4010,4011,4012,4013,4014,4015,4016,4017,4018,4019,4020,4021,4022,4023,4024,4025,4026,4027,4028,4029,4030,4031,4032,4033,4034,4035,4036,4037,4038,4039,4040,4041,4042,4043,4044,4045,4046,4047,4048,4049,4050,4051,4052,4053,4054,4055,4056,4057,4058,4059,4060,4061,4062,4063,4064,4065,4066,4067,4068,4069,4070,4071,4072,4073,4074,4075,4076,4077,4078,4079,4080,4081,4082,4083,4084,4085,4086,4087,4088,4089,4090,4091,4092,4093,4094,4095,4096,4097,4098,4099,4100,4101,4102,4103,4104,4105,4106,4107,4108,4109,4110,4111,4112,4113,4114,4115,4116,4117,4118,4119,4120,4121,4122,4123,4124,4125,4126,4127,4128,4129,4130,4131,4132,4133,4134,4135,4136,4137,4138,4139,4140,4141,4142,4143,4144,4145,4146,4147,4148,4149,4150,4151,4152,4153,4154,4155,4156,4157,4158,4159,4160,4161,4162,4163,4164,4165,4166,4167,4168,4169,4170,4171,4172,4173,4174,4175,4176,4177,4178,4179,4180,4181,4182,4183,4184,4185,4186,4187,4188,4189,4190,4191,4192,4193,4194,4195,4196,4197,4198,4199,4200,4201,4202,4203,4204,4205,4206,4207,4208,4209,4210,4211,4212,4213,4214,4215,4216,4217,4218,4219,4220,4221,4222,4223,4224,4225,4226,4227,4228,4229,4230,4231,4232,4233,4234,4235,4236,4237,4238,4239,4240,4241,4242,4243,4244,4245,4246,4247,4248,4249,4250,4251,4252,4253,4254,4255,4256,4257,4258,4259,4260,4261,4262,4263,4264,4265,4266,4267,4268,4269,4270,4271,4272,4273,4274,4275,4276,4277,4278,4279,4280,4281,4282,4283,4284,4285,4286,4287,4288,4289,4290,4291,4292,4293,4294,4295,4296,4297,4298,4299,4300,4301,4302,4303,4304,4305,4306,4307,4308,4309,4310,4311,4312,4313,4314,4315,4316,4317,4318,4319,4320,4321,4322,4323,4324,4325,4326,4327,4328,4329,4330,4331,4332,4333,4334,4335,4336,4337,4338,4339,4340,4341,4342,4343,4344,4345,4346,4347,4348,4349,4350,4351,4352,4353,4354,4355,4356,4357,4358,4359,4360,4361,4362,4363,4364,4365,4366,4367,4368,4369,4370,4371,4372,4373,4374,4375,4376,4377,4378,4379,4380,4381,4382,4383,4384,4385,4386,4387,4388,4389,4390,4391,4392,4393,4394,4395,4396,4397,4398,4399,4400,4401,4402,4403,4404,4405,4406,4407,4408,4409,4410,4411,4412,4413,4414,4415,4416,4417,4418,4419,4420,4421,4422,4423,4424,4425,4426,4427,4428,4429,4430,4431,4432,4433,4434,4435,4436,4437,4438,4439,4440,4441,4442,4443,4444,4445,4446,4447,4448,4449,4450,4451,4452,4453,4454,4455,4456,4457,4458,4459,4460,4461,4462,4463,4464,4465,4466,4467,4468,4469,4470,4471,4472,4473,4474,4475,4476,4477,4478,4479,4480,4481,4482,4483,4484,4485,4486,4487,4488,4489,4490,4491,4492,4493,4494,4495,4496,4497,4498,4499,4500,4501,4502,4503,4504,4505,4506,4507,4508,4509,4510,4511,4512,4513,4514,4515,4516,4517,4518,4519,4520,4521,4522,4523,4524,4525,4526,4527,4528,4529,4530,4531,4532,4533,4534,4535,4536,4537,4538,4539,4540,4541,4542,4543,4544,4545,4546,4547,4548,4549,4550,4551,4552,4553,4554,4555,4556,4557,4558,4559,4560,4561,4562,4563,4564,4565,4566,4567,4568,4569,4570,4571,4572,4573,4574,4575,4576,4577,4578,4579,4580,4581,4582,4583,4584,4585,4586,4587,4588,4589,4590,4591,4592,4593,4594,4595,4596,4597,4598,4599,4600,4601,4602,4603,4604,4605,4606,4607,4608,4609,4610,4611,4612,4613,4614,4615,4616,4617,4618,4619,4620,4621,4622,4623,4624,4625,4626,4627,4628,4629,4630,4631,4632,4633,4634,4635,4636,4637,4638,4639,4640,4641,4642,4643,4644,4645,4646,4647,4648,4649,4650,4651,4652,4653,4654,4655,4656,4657,4658,4659,4660,4661,4662,4663,4664,4665,4666,4667,4668,4669,4670,4671,4672,4673,4674,4675,4676,4677,4678,4679,4680,4681,4682,4683,4684,4685,4686,4687,4688,4689,4690,4691,4692,4693,4694,4695,4696,4697,4698,4699,4700,4701,4702,4703,4704,4
```

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
 Virology 188, 102-113, 1992  
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
 A:Reference number: A40244; MUID:92230206; PMID:1314449  
 A:Accession: A40244  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3010 <CHE>  
 A:Cross-references: GB:M84754  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPW>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,233,234,250,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 98.0%; Score 48; DB 1; Length 3010;  
 Best Local Similarity 90.9%; Pred. No. 2.4;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11  
 |||||  
 Db 1678 GSVVIVGRIVL 1688

RESULT 6  
 PS0326  
 polyprotein - hepatitis C virus (isolate Fla) (fragments)  
 C:Species: hepatitis C virus  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000  
 C:Accession: PS0326  
 R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.  
 Gene 105, 167-172, 1991  
 A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype  
 A:Reference number: PS0326; MUID:92039028; PMID:1718820  
 A:Accession: PS0326  
 A:Molecule type: genomic RNA  
 A:Residues: 1-492 <LIJ>  
 A:Cross-references: GB:M60220  
 A:Note: this sequence corresponds to nonstructural protein NS3 region  
 A:Note: translation of the nucleotide sequence is not complete  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: polyprotein

Query Match 89.8%; Score 44; DB 2; Length 492;  
 Best Local Similarity 90.9%; Pred. No. 2.3;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11  
 |||||  
 Db 229 GSVVIVGRIVL 239

RESULT 7  
 PC2219  
 polypeptide - hepatitis C virus (type 5a) (fragments)  
 N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote  
 C:Species: hepatitis C virus  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Nov-2000  
 C:Accession: PC2219  
 R:Stuyver, L.; Arnhen, W.V.; Wyseur, A.; Maertens, G.  
 Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
 A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the h  
 A:Reference number: PC2219; MUID:94338342; PMID:7520237  
 A:Accession: PC2219

A:Molecule type: mRNA  
 A:Residues: 1-876 <STU>  
 A:Cross-references: GB:L29577; GB:L29578; GB:L29579  
 A:Experimental source: serum  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: glycoprotein  
 F:1-191/Product: core #status predicted <COE>  
 F:68-78/Region: variable  
 F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
 F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
 F:248-338/Region: E2  
 F:339-411/Region: NS1 (amino end)  
 F:412-783/Product: NS3 #status predicted <NSR>  
 F:784-837/Product: NS4A #status predicted <NSA>  
 F:838-876/Product: NS4B #status predicted <NSB>  
 F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 89.8%; Score 44; DB 2; Length 876;  
 Best Local Similarity 81.8%; Pred. No. 4;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11  
 |||||  
 Db 804 GSVVIVGRIVL 814

RESULT 8  
 GNVVCH  
 genome polyprotein - hepatitis C virus (strain H)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) (nonstr  
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C:Accession: A36814; A41546  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 submitted to Genbank, July 1992  
 A:Description: Genomic structure of the human prototype strain H of hepatitis C viru  
 A:Reference number: A36814  
 A:Accession: A36814  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <INC>  
 A:Cross-references: GB:M67463; NID:9329737; PIDN:AAA45534.1; PID:G329738  
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: com  
 A:Reference number: A41546; MUID:92052256; PMID:1658800  
 A:Contents: annotation  
 A:Note: neither amino acid nor nucleotide sequence is given  
 C:Superfamily: hepatitis C virus genome polyprotein  
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruc  
 F:1-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPW>  
 F:192-389/Product: major envelope protein E #status predicted <NEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,224

Query Match 89.8%; Score 44; DB 1; Length 3011;  
 Best Local Similarity 90.9%; Pred. No. 13;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11  
 |||||  
 Db 1678 GSVVIVGRIVL 1688

```

RESULT 9
S4070
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:DI0749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Experimental source: isolate HC-J1
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:116-191/Product: capsid protein C #status predicted <CPC>
F:192-389/Product: envelope protein M #status predicted <EPM>
F:390-729/Product: major envelope protein E #status predicted <MEE>
F:1007-1615/Product: nonstructural protein NS1 #status predicted <NS1>
F:1230-1237/Product: nonstructural protein NS2 #status predicted <NS2>
F:1312-1319/Product: nonstructural protein NS3 #status predicted <NS3>
F:1316-1319/Product: nonstructural protein NS4 #status predicted <NS4>
F:1616-1862/Product: nonstructural protein NS5 #status predicted <NS5>
F:1863-2013/Product: nonstructural protein NS6 #status predicted <NS6>
F:2014-3011/Product: nonstructural protein NS7 #status predicted <NS7>
Query Match 89.8%; Score 44; DB 1; Length 3011;
Best Local Similarity 90.9%; Pred. No. 13;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRVIL 11
| | | | | | | | | |
Db 1678 GCVVIVGRVIL 1688

RESULT 10
JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
A:Reference number: JC5620; MUID:97366593; PMID:9223423
A:Accession: JC5620
A:Molecule type: mRNA
A:Residues: 1-3014 <GB>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: The translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:384-408/Product: hypervariable #status predicted

```

```

F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitis C virus NS4b; nonstructural protein NS5
F:1231-1238/Product: nucleotide-binding motif A (P-loop)
F:1313-1318/Product: nucleotide-binding motif B
F:1317-1320/Product: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1864-2014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Product: interferon sensitivity determining #status predicted

Query Match 89.8%; Score 44; DB 1; Length 3014;
Best Local Similarity 81.8%; Pred. No. 13;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRVIL 11
| | | | | | | | | |
Db 1679 GSAIVGRVIL 1689

RESULT 11
GNWVC3
genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PQ0403; PQ0404
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; MUID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA45676.1; PID:g329874
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:DJ10128
A:Experimental source: isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Product: hepatitis C virus NS4a #status predicted <NS4a>
F:1316-1319/Product: nucleotide-binding motif A (P-loop)
F:1316-1319/Product: DEXH motif
F:1616-1862/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,334,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

```

```

Query Match 87.8%; Score 43; DB 1; Length 3011;
Best Local Similarity 81.8%; Pred. No. 20;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSVVIVGRVIL 11
| | | | | | | | | |
Db 1678 GCVVIVGRVIL 1688

```

## RESULT 12

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 17-Nov-2000

C:Accession: JQ1366

R:Kremendorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 &lt;KRE&gt;

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 85.7%; Score 42; DB 2; Length 716;

Best Local Similarity 72.7%; Pred. No. 7.6;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11

Db 627 GCVVIVGRVIL 637

## RESULT 13

G70978

Probable copper-transporting ATPase 11/9 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 18-Aug-2000

C:Accession: G70978

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70978

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-718 &lt;COL&gt;

A:Cross-references: GB:292771; GB:AL123456; NID:93242259; PIDN:CAB07083.1; PID:gl877325

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: ctpC

C:Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding d

F:128-434/Domain: ATPase transduction domain homology &lt;ATT&gt;

F:532-676/Domain: ATPase nucleotide-binding domain homology &lt;ATN&gt;

Query Match

Best Local Similarity 75.5%; Score 37; DB 2; Length 718;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRIVL 11

Db 286 GSVVIVGRVV 296

## RESULT 14

T41395

Probable dna polymerase alpha-primase associated subunit - fission yeast (Schizosacchar

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T41395

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, May 1998

A:Reference number: 221991

A:Accession: T41395

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-574 &lt;WOO&gt;

A:Cross-references: EMBL:AL023704; PIDN:CAAL9261.1; GSPDB:GN00068; SPDB:SPCC553.09c

A:Experimental source: strain 972h-; cosmid c553

C:Genetics:

A:Gene: SPOB:SPCC553.09c

A:Map position: 3

A:Introns: 89/2; 415/3; 518/3

Query Match 71.4%; Score 35; DB 2; Length 574;

Best Local Similarity 77.8%; Pred. No. 1.2e+02;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 VVIVGRIVL 11

Db 205 VVVVGRIVV 213

## RESULT 15

E89949

valine-tRNA ligase [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: E89949

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; C

ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, T

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E89949

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-876 &lt;KUR&gt;

A:Cross-references: GB:BA000018; PID:gl3701460; PIDN:BAB42754.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: valS

C:Superfamily: valine-tRNA ligase

Query Match

Best Local Similarity 71.4%; Score 35; DB 2; Length 876;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SVVIVGRIVL 11

Db 794 SVVIAGKVV 803

Search completed: August 30, 2003, 19:20:34

Job time : 2.90532 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 18:01:52 ; Search time 0.544792 Seconds  
(without alignments)  
949.524 Million cell updates/sec

Title: US-09-965-594-26  
Perfect score: 49  
Sequence: 1 GSVVINGRVL 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	3010	1	POLG_HCVBK
2	48	98.0	3010	1	POLG_HCVJA
3	48	98.0	3010	1	POLG_HCVJT
4	48	98.0	3010	1	POLG_HCVTW
5	44	89.8	3011	1	POLG_HCVH
6	43	87.8	3011	1	POLG_HCVI
7	37	75.5	718	1	CTPC_MYCTU
8	35	71.4	574	1	DPO2_SCHPO
9	34	69.4	196	1	YM07_YEAST
10	34	69.4	260	1	MCH_METTR
11	34	69.4	859	1	ENV_EIAV1
12	34	69.4	859	1	ENV_EIAV2
13	34	69.4	859	1	ENV_EIAV3
14	34	69.4	859	1	ENV_EIAV9
15	34	69.4	859	1	ENV_EIAVC
16	34	69.4	859	1	ENV_EIAVM
17	34	69.4	859	1	ENV_EIAVY
18	34	69.4	860	1	ENV_EIAV5
19	33	67.3	316	1	MCH_ARCFU
20	33	67.3	405	1	ASSY_PSEAE
21	33	67.3	456	1	SHU7_ECOLI
22	33	67.3	457	1	CYSG_ECOLI
23	33	67.3	560	1	J1160_HORVU
24	33	67.3	725	1	CTPC_MYCLE
25	33	67.3	968	1	MLL2_MYCTU
26	33	67.3	1158	1	CND1_SCHPO
27	32	65.3	249	1	COBM_RHOER
28	32	65.3	332	1	PLSX_THETN
29	32	65.3	388	1	Y8C8_SALTU
30	32	65.3	411	1	DBOB_LACLA
31	32	65.3	411	1	DBOB_LACLC
32	32	65.3	417	1	HS47_HUMAN
33	32	65.3	417	1	HS47_MOUSE

34 32 65.3 417 1 HS47\_RAT  
35 32 65.3 418 1 CBP2\_HUMAN  
36 32 65.3 509 1 YFCC\_HAEIN  
37 32 65.3 752 1 CTPB\_MYCTU  
38 32 65.3 855 1 YB29\_YEAST  
39 31 63.3 84 1 RL27\_CHLTE  
40 31 63.3 193 1 YNES\_BACSU  
41 31 63.3 223 1 VTAY\_LAMBD  
42 31 63.3 228 1 NEUA\_HAEIN  
43 31 63.3 231 1 RS2\_SULSO  
44 31 63.3 251 1 FGFN\_RAT  
45 31 63.3 258 1 UPKA\_BOVIN

## ALIGNMENTS

RESULT 1  
POLG\_HCVBK  
ID POLG\_HCVBK STANDARD; PRT; 3010 AA.  
AC P26653;  
DT 01-AUG-1992 (Rel. 23, Created)  
DI 01-AUG-1992 (Rel. 23, Last sequence update)  
DI 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P36); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DE Hepatitis C virus (isolate BK) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
OC Hepacivirus.  
OX NCBI\_TaxID=11105;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1847440;  
RA Takamizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;  
RT "Structure and organization of the hepatitis C virus genome isolated from human carriers";  
RL J. Virol. 65:1105-1113(1991).  
[2]  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=96235224; PubMed=8647104;  
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetky L.;  
RT "Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase";  
RL Eur. J. Biochem. 237:611-618(1996).  
[3]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.  
RX MEDLINE=97015088; PubMed=8861916;  
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka M., Moomaw E.W., Adachi T., Hostomsky Z.;  
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site";  
RL Cell 87:331-342(1996).  
[4]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.  
RX MEDLINE=98227846; PubMed=9568891;  
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;  
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2-A resolution structure in a hexagonal crystal form";  
RL Protein Sci. 7:837-847(1998).  
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +



(RNA)(N).  
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 SIMILAR C AND MRNA.  
 -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 -----  
 This Swiss-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 EMBL: M58335; AAA72945.1; --  
 PIR: A38465; GNVVTC.  
 PDB: 1A10; 25-MAR-98.  
 PDB: 1JXP; 14-JAN-98.  
 PDB: 1NS3; 08-APR-98.  
 PDB: 1C2P; 15-NOV-00.  
 PDB: 1CSJ; 08-NOV-99.  
 PDB: 1GX5; 09-APR-02.  
 PDB: 1GX6; 10-APR-02.  
 PDB: 1QV6; 26-JUN-00.  
 PDB: 80HM; 20-APR-99.  
 MEROPS: S29.001; --  
 MEROPS: U39.001; --  
 InterPro: IPR001410; DEAD.  
 InterPro: IPR002522; HCV\_capsid.  
 InterPro: IPR002521; HCV\_core.  
 InterPro: IPR002519; HCV\_env.  
 InterPro: IPR002531; HCV\_NS1.  
 InterPro: IPR002518; HCV\_NS2.  
 InterPro: IPR004109; HCV\_NS3.  
 InterPro: IPR000745; HCV\_NS4a.  
 InterPro: IPR001490; HCV\_NS4b.  
 InterPro: IPR002868; HCV\_NS5a.  
 InterPro: IPR002166; HCV\_RdRP.  
 InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 InterPro: IPR007094; RNA\_pol\_PSVir.  
 Pfam: PF01543; HCV\_capsid; 1.  
 Pfam: PF01542; HCV\_core; 1.  
 Pfam: PF01539; HCV\_env; 1.  
 Pfam: PF01560; HCV\_NS1; 1.  
 Pfam: PF01538; HCV\_NS2; 1.  
 Pfam: PF02907; HCV\_NS3; 1.  
 Pfam: PF01006; HCV\_NS4a; 1.  
 Pfam: PF01001; HCV\_NS4b; 1.  
 Pfam: PF01506; HCV\_NS5a; 1.  
 Pfam: PF00998; Viral\_RdRP; 1.  
 ProDom: PD186062; HCV\_NS1; 1.  
 SMART: SM00487; DEXDC; 1.  
 PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 3D-structure.  
 INIT\_MET 1 1  
 CHAIN 1 115  
 CHAIN 116 191  
 CHAIN 192 383  
 CHAIN 384 729  
 CHAIN 730 1006  
 CHAIN 1007 1615  
 CHAIN 1616 1862  
 CHAIN 1863 2013  
 CHAIN 2014 3010  
 CHAIN 347 369  
 TRANSMEM 1083 1083  
 ACT\_SITE 1107 1107  
 ACT\_SITE 1165 1165  
 ACT\_SITE 1230 1237  
 NP\_BIND (POTENTIAL).  
 CAPSID PROTEIN C (POTENTIAL).  
 MATRIX PROTEIN (POTENTIAL).  
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 PROTEASE/HELICASE NS3 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 POTENTIAL.  
 CHARGE RELAY SYSTEM.  
 CHARGE RELAY SYSTEM.  
 CHARGE RELAY SYSTEM.  
 ATP (POTENTIAL).

FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 FT STRAND 1031 1035  
 FT HELIX 1039 1047  
 FT STRAND 1050 1050  
 FT STRAND 1059 1063  
 FT STRAND 1068 1074  
 FT TURN 1075 1076  
 FT STRAND 1077 1081  
 FT HELIX 1082 1085  
 FT TURN 1086 1087  
 FT STRAND 1090 1092  
 FT TURN 1093 1094  
 FT STRAND 1095 1097  
 FT STRAND 1101 1103  
 FT TURN 1104 1107  
 FT STRAND 1108 1112  
 FT STRAND 1120 1120  
 FT STRAND 1122 1122  
 FT STRAND 1129 1133  
 FT TURN 1135 1136  
 FT STRAND 1139 1144  
 FT STRAND 1149 1157  
 FT HELIX 1158 1161  
 FT TURN 1162 1163  
 FT TURN 1165 1166  
 FT STRAND 1168 1171  
 FT TURN 1172 1174  
 FT STRAND 1175 1186  
 FT TURN 1187 1188  
 FT STRAND 1189 1197  
 FT HELIX 1198 1202  
 FT TURN 1203 1204  
 FT STRAND 1680 1688  
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;

Query Match 98.0%; Score 48; DB 1; Length 3010;

Best Local Similarity 90.9%; Pred. No. 1.1;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVIVGRVL 11

Db 1678 GSVIVGRIL 1688

RESULT 2

POLG\_HCVJA STANDARD; PRT: 3010 AA.

ID POLG\_HCVJA AC P26662;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Japanese) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=9108550; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from  
 RT Japanese patients with non-A, non-B hepatitis.";   
 EL proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP DISCUSSION OF SEQUENCE.  
 RA MEDLINE=91192160; PubMed=1849489;  
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraishi K.,  
 RA Ohkoshi S., Shimotohno K.;  
 RT "Molecular structure of the Japanese hepatitis C viral genome.";   
 RL FEBS Lett. 280:325-328(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC [RNA](N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DE EMBL; D90208; BAA14233.1; -;  
 DE PIR; A39253; GNWVCJ.  
 DE HSSP; P26663; LUXP.  
 DE MEROPS; S29.001; -;  
 DE MEROPS; U39.001; -;  
 DE InterPro; IPR001410; DEAD.  
 DE InterPro; IPR002522; HCV\_capsid.  
 DE InterPro; IPR002521; HCV\_core.  
 DE InterPro; IPR002519; HCV\_env.  
 DE InterPro; IPR002531; HCV\_NS1.  
 DE InterPro; IPR002518; HCV\_NS2.  
 DE InterPro; IPR004109; HCV\_NS3.  
 DE InterPro; IPR000745; HCV\_NS4a.  
 DE InterPro; IPR001490; HCV\_NS4b.  
 DE InterPro; IPR002868; HCV\_NS5a.  
 DE InterPro; IPR002166; HCV\_NS5b.  
 DE InterPro; IPR001650; Helicase\_C.  
 DE InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DE InterPro; IPR007094; RNA\_pol\_PSVir.  
 DE Pfam; PF01543; HCV\_capsid; 1.  
 DE Pfam; PF01542; HCV\_core; 1.  
 DE Pfam; PF01539; HCV\_env; 1.  
 DE Pfam; PF01560; HCV\_NS1; 1.  
 DE Pfam; PF01538; HCV\_NS2; 1.  
 DE Pfam; PF02907; HCV\_NS3; 1.  
 DE Pfam; PF01006; HCV\_NS4a; 1.  
 DE Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT CHAIN 3011 369  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2788 2788  
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 98.0%; Score 48; DB 1; Length 3010;  
 Best Local Similarity 90.9%; Pred. No. 1.1;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSVVIVGRIVL 11

Db 1678 GSVVIVGRIVL 1688

RESULT 3

POLG\_HCVJT

ID POLG\_HCVJT STANDARD; PRT: 3010 AA.

AC Q00269;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate HC-JT) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=31642;

RN [1]

SEQUENCE FROM N.A.  
 MEDLINE-92295714: PubMed-1118627:  
 Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
 Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
 "Molecular cloning of hepatitis C virus genome from a single Japanese  
 carrier: sequence variation within the same individual and among  
 infected individuals.";  
 Virus Res. 23:39-53(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D11168; BAA01943.1; -  
 CC PIR: A45573; A45573.  
 CC PDB: 1A10; 25-MAR-98.  
 CC PDB: 1JXP; 14-JAN-98.  
 CC MEROPS: S29.001; -  
 CC MEROPS: U39.001; -  
 CC  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR002522; HCV\_capsid.  
 CC InterPro: IPR002321; HCV\_core.  
 CC InterPro: IPR002531; HCV\_gn1.  
 CC InterPro: IPR002531; HCV\_gn1.  
 CC InterPro: IPR002518; HCV\_NS2.  
 CC InterPro: IPR004109; HCV\_NS3.  
 CC InterPro: IPR000745; HCV\_NS4a.  
 CC InterPro: IPR001490; HCV\_NS4b.  
 CC InterPro: IPR002868; HCV\_NS5a.  
 CC InterPro: IPR002166; HCV\_RGRP.  
 CC InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro: IPR007094; RNA\_pol\_Psvir.  
 CC Pfam: PF01543; HCV\_core; 1.  
 CC Pfam: PF01542; HCV\_core; 1.  
 CC Pfam: PF01539; HCV\_gn1; 1.  
 CC Pfam: PF01560; HCV\_gn1; 1.  
 CC Pfam: PF01538; HCV\_NS2; 1.  
 CC Pfam: PF02907; HCV\_NS3; 1.  
 CC Pfam: PF01006; HCV\_NS4a; 1.  
 CC Pfam: PF01001; HCV\_NS4b; 1.  
 CC Pfam: PF01506; HCV\_NS5a; 1.  
 CC Pfam: PF00271; helicase; 1.  
 CC Pfam: PF00998; Viral\_RdRp; 1.  
 CC ProDom: PD186062; HCV\_NS1; 1.  
 CC SMART: SM00487; DEADc; 1.  
 CC  
 CC Core protein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 CC Polypeptide; Coat protein; Envelope protein; Helicase; ATP-binding;  
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 CC 3D-structure.  
 CC INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 CC CELLULAR AMINOPEPTIDASE.  
 CC CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).  
 CC CHAIN 116 191 MATRIX ENVELOPE (POTENTIAL).  
 CC CHAIN 192 393 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 CC CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 CC CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 FT TRANSMEM 347 369 POTENTIAL.  
 FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT NP\_BIND 1230 1237 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;  
 Query Match 98.0%; Score 48; DB 1; Length 3010;  
 Best Local Similarity 90.9%; Pred. No. 1.1;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSVWIVGRVL 11  
 Db 1678 GSVWIVGRIL 1688  
 RESULT 4  
 POLG\_HCVTW STANDARD; PRT: 3010 AA.  
 ID POLG\_HCVTW  
 AC P29846;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI\_TaxID=31645;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92230206; PubMed-1314449;  
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RT "The Taiwanese hepatitis C virus genome: sequence determination and  
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the p6  
 CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M84754; .. NOT\_ANNOTATED\_CDS.  
 DR PIR: A40244; GNRVTV.  
 DR PDB: IN63; 25-FEB-03.  
 DR PDB: INS3; 08-APR-98.  
 DR MEROPS: S29.001; ..  
 DR MEROPS: S29.001; ..  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002522; HCV\_capsid.  
 DR InterPro: IPR002521; HCV\_core.  
 DR InterPro: IPR002519; HCV\_env.  
 DR InterPro: IPR002531; HCV\_NS1.  
 DR InterPro: IPR002518; HCV\_NS2.  
 DR InterPro: IPR004109; HCV\_NS3.  
 DR InterPro: IPR000745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR002868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_NS5b.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PS\_vir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF02907; HCV\_NS3; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00098; Viral\_RDRP; 1.  
 DR ProDom: PD186062; HCV\_NS1; 1.  
 DR SMART: SM00487; DEXdc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 FT CELLULAR AMINOPEPTIDASE.  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA: 327047 MW: 5426755CDPE215 CRC64;

Query Match 98.0%; Score 48; DB 1; Length 3010;  
 Best Local Similarity 90.9%; Pred. No. 1.1;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSVVIGRVL 11  
 |||||  
 Db 1678 GSVVIGRIL 1688

## RESULT 5

POLG\_HCVH STANDARD; PRT: 3011 AA.  
 ID POLG\_HCVH  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT \*Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT \*Structure of the hepatitis C virus RNA helicase domain.\*;  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT \*Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 RT of unwinding.\*;  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.



FT TURN 1598 1598  
 FT HELIX 1606 1611  
 FT TURN 1614 1618  
 FT STRAND 1622 1623  
 FT STRAND 1627 1627  
 FT STRAND 1635 1636  
 FT HELIX 1640 1652  
 SQ SEQUENCE 3011 AA: 772CBB29CCD94753 CRC64;

Query Match 89.8%; Score 44; DB 1; Length 3011;  
 Best Local Similarity 90.9%; Pred. No. 5.7;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSWIVIGRVL 11  
 | | | | | | | | | |  
 DB 1678 GCWIVIGRVL 1698

RESULT 6  
 POLG\_HCV1 STANDARD; PRT: 3011 AA.  
 AC P26664;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (P35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate 1) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11104;  
 RN [1]  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,  
 RA Bradley D.W., Kuo G., Houghton M.;  
 RT "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; M62321; AAA45676.1; -  
 CC PIR; A30166; GNWVC3.  
 CC PDB; 1A1V; 16-FEB-99.  
 CC MEROPS; S29.001; -  
 CC MEROPS; U39.001; -  
 DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR004109; HCV\_NS3.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_Ndrp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR007095; RNA\_pol\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_Rdrp; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PolyProtein; Glycoprotein; Transferrin; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 SQ SEQUENCE 3011 AA: 327197 MW: 65F8C9447FCE5AF9 CRC64;  
 Query Match 87.8%; Score 43; DB 1; Length 3011;  
 Best Local Similarity 81.8%; Pred. No. 8.7;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 GSVWVGRVIL 11
DB 1678 GCVVGRVIL 1688

RESULT 7
CTPC_MYCTU STANDARD: PRT: 718 AA.
AC P96875:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable cation-transporting P-type ATPase C (EC 3.6.3.-).
GN CTPC OR RV3270 OR MT3370 OR MYCY71.10.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekait F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RA STRAIN=CDC 1551 / Oshkosh;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Z92771; CAB07083.1;
CC EMBL: AE007146; AKA47711.1;
CC F1R: G70978; G70978.
CC TIGR: MT3370;
CC TubercuList: Rv3270;
CC InterPro: IPR006416; ATPase-IB_hvy.
CC InterPro: IPR001757; ATPase-El-E2.
CC InterPro: IPR006404; Heavy_metal_ATPase.
CC InterPro: IPR005834; Hydrolyase.
CC Pfam: PF00122; El-E2_ATPase; 1.
CC Pfam: PF00702; Hydrolyase; 1.
CC PRINTS: PR00119; CATATPASE.
CC TIGRfams: TIGR01512; ATPase-IB2_Cd; 1.
CC TIGRfams: TIGR01525; ATPase-IB_hvy; 1.
CC TIGRfams: TIGR01494; ATPase_P-type; 3.
CC PROSITE: PS00154; ATPASE_E1_E2; 1.

```

Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;

KW Complete proteome. 194 POTENTIAL.

KW TRANSMEM 174 380 POTENTIAL.

FT TRANSMEM 360 636 POTENTIAL.

FT TRANSMEM 516 636 POTENTIAL.

FT TRANSMEM 577 697 POTENTIAL.

FT MOD\_RES 408 408 PHOSPHORYLATION (BY SIMILARITY).

FT METAL 610 610 MAGNESIUM (BY SIMILARITY).

FT METAL 614 614 MAGNESIUM (BY SIMILARITY).

FT SEQUENCE 718 AA: 76495 MW: 85D6C93AFE636315 CRC64;

Query Match 75.5%; Score 37; DB 1; Length 718;

Best Local Similarity 63.6%; Pred. NO. 30;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVWVGRVIL 11

DB 286 GSVWVGRVIL 296

RESULT 8

DPO2\_SCHPO STANDARD: PRT: 574 AA.

ID DPO2\_SCHPO

AC 074946; 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable DNA polymerase alpha subunit B.

GN SPCC553.09C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA MEDLINE=21848401; PubMed=11859360;

RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeall C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Skelton J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,

RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe";

RL Nature 415:871-880(2002).

CC -!- FUNCTION: MAY PLAY AN ESSENTIAL ROLE AT THE EARLY STAGE OF

CC CHROMOSOMAL DNA REPLICATION BY COUPLING THE POLYMERASE

CC ALPHA/PRIMASE COMPLEX TO THE CELLULAR REPLICATION MACHINERY (BY

CC SIMILARITY).

CC -!- SUBUNIT: DNA POLYMERASE ALPHA-PRIMASE IS A FOUR SUBUNIT ENZYME

CC (SUBUNITS A, B, C AND D), WHICH IS ASSEMBLED THROUGHOUT THE CELL

CC CYCLE. THE LARGEST SUBUNIT (SUBUNIT A) HAS DNA POLYMERASE

CC ACTIVITY, THE TWO SMALLEST SUBUNITS (SUBUNITS C AND D) HAVE DNA

CC PRIMASE ACTIVITY. SUBUNIT B BINDS TO SUBUNIT A.



CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER (BY  
 CC SIMILARITY). BELONGS TO THE DNA POLYMERASE ALPHA SUBUNIT B FAMILY.  
 CC  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL023704; CAA19261.1; -.  
 CC PIR; T41395; T41395.  
 CC GeneDB\_Spombe; SPCC553.09c; -.  
 CC pfam; PF04058; DNA\_pol\_alpha\_B; 1.  
 CC DNA replication; Nuclear protein; Phosphorylation.  
 CC  
 CC QUERY MATCH 71.4%; Score 35; DB 1; Length 574;  
 CC Best Local Similarity 77.8%; Pred. No. 58;  
 CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 3 VVIVGRIVL 11  
 CC ||:|||||  
 CC DB 205 VVVVGRIVV 213  
 CC  
 CC RESULT 9  
 CC ID Y07\_YEAST STANDARD; PRT; 196 AA.  
 CC AC Q04487;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Putative succinate dehydrogenase cytochrome B subunit, mitochondrial  
 CC precursor.  
 CC GN YMR118C OR YMR9718.17C.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC OX NCBI\_TaxID:4932;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-S288c / AB972;  
 CC RX PubMed:9169872;  
 CC RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 CC Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 CC Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 CC Rice P., Skelton J., Walsh S., Whitehead S., Barrrell B.G.;  
 CC "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 CC XIII.";  
 CC RL Nature 387:90-93(1997).  
 CC  
 CC -!- FUNCTION: MONO-HEME CYTOCHROME B. INVOLVED IN SYSTEM II OF THE  
 CC MITOCHONDRIAL ELECTRON TRANSPORT CHAIN WHICH IS RESPONSIBLE FOR  
 CC TRANSFERRING ELECTRONS FROM SUCCINATE TO UBIQUINONE (COENZYME Q)  
 CC (BY SIMILARITY).  
 CC  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 CC inner membrane (By similarity).  
 CC  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B560 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z49702; CAA89756.1; -.  
 CC PIR; S54580; S54580.  
 CC SGD; S0004724; YMR118C.

DR InterPro; IPR000701; sdh\_cyt.  
 DR Pfam; PF01127; Sdh\_cyt; 1.  
 DR PROSITE; PS01000; SDH\_CYT\_1; 1.  
 DR PROSITE; PS01001; SDH\_CYT\_2; 1.  
 KW Hypothetical protein; Tricarboxylic acid cycle; Electron transport;  
 KW Heme; Transmembrane; Mitochondrion; Transit peptide.  
 FT TRANSIT ? 196 MITOCHONDRION (POTENTIAL).  
 FT CHAIN ? 196 PUTATIVE SUCCINATE DEHYDROGENASE  
 FT CYTOCHROME B SUBUNIT.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 175 195 POTENTIAL.  
 SQ SEQUENCE 196 AA; 22309 MW; 414139989B2B057 CRC64;  
 Query Match 69.4%; Score 34; DB 1; Length 196;  
 Best Local Similarity 54.5%; Pred. No. 34;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GSVVIVGRIVL 11  
 DB 179 GSVVIVGRIVL 189  
 RESULT 10  
 MCH\_METTR STANDARD; PRT; 260 AA.  
 ID MCH\_METTR  
 AC Q9RPD4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE N(5),N(1b)-methenyltetrahydromethanopterin cyclohydrolase  
 DE (EC 3.5.4.27) (Methenyl-HMPT cyclohydrolase) (Fragment).  
 GN MCH.  
 OS Methylosinus trichosporium.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Methylocystaceae; Methylosinus.  
 OX NCBI\_TaxID:426;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OB3b;  
 RX MEDLINE-99412275; PubMed-10482517;  
 RA Vorholt J.A., Chistosterdova L.V., Stolyar S.M., Thauer R.K.,  
 RA Lidstrom M.E.;  
 RT "Distribution of tetrahydromethanopterin-dependent enzymes in  
 RT methylophilic bacteria and phylogeny of methenyl  
 RT tetrahydromethanopterin cyclohydrolases.";  
 RL J. Bacteriol. 181:5750-5757(1999).  
 CC -!- FUNCTION: Catalyzes the hydrolysis of methenyl-H(4)MPT to N(5)-  
 CC formyl-H(4)MPT (By similarity).  
 CC -!- CATALYTIC ACTIVITY: 5,10-methenyl-5,6,7,8-tetrahydromethanopterin  
 CC + H(2)O = N(5)-formyl-5,6,7,8-tetrahydromethanopterin.  
 CC -!- PATHWAY: H(4)MPT-dependent pathway of formaldehyde oxidation;  
 CC third step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE MCH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF162786; AAD56174.1; -.  
 CC HSP; P94954; IQLM.  
 DR HAMAP; MF\_00486; -; 1.  
 DR InterPro; IPR003209; Cyclohydrolase.  
 DR Pfam; PF02289; MCH; 1.  
 DR ProDom; PD011637; Cyclohydrolase; 1.  
 KW Hydrolase; One-carbon metabolism.  
 FT NON\_TER 1  
 SQ SEQUENCE 260 AA; 27622 MW; BD647C0DBF03C6A8 CRC64;





FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 859 AA; 97188 MW; D86E4E171E39B32 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;

Best Local Similarity 77.8%; Pred. No. 1.3e-02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRI 9

Db 801 GLVIVGRI 809

RESULT 13

ENV\_EIAV3

ID ENV\_EIAV3 STANDARD; PRT; 859 AA.

AC P22429;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein

GP90; Coat protein GP45].

GN ENV.

OS Equine infectious anemia virus (Clone P3.2-3) (EIAV).

OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11668;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=8807207; PubMed=2825406;

RA Payne S.L., Fang F.D., Liu C.P., Dhruva B.R., Rwambo P., Issel C.J.,

RA Montelaro R.C.;

RT "Antigenic variation and lentivirus persistence: variations in

envelope gene sequences during EIAV infection resemble changes

RT reported for sequential isolates of HIV.\*;

RL Virology 161:321-331(1987).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M18387; AAA66409.1; -

DR PIR; C34027; VCLJJE3.

DR InterPro: IPR001027; GP45\_EIAV.

DR InterPro: IPR001361; GP90\_EIAV.

DR Pfam; PF01045; EIAV\_GP45; 1.

DR Pfam; PF00971; EIAV\_GP90; 1.

KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 859 ENV POLYPROTEIN.

FT CHAIN 23 444 COAT PROTEIN GP90.

FT CHAIN 445 859 COAT PROTEIN GP45.

FT CHAIN 75 93 POTENTIAL.

FT TRANSMEM 446 462 POTENTIAL.

FT TRANSMEM 614 636 POTENTIAL.

FT TRANSMEM 787 807 POTENTIAL.

FT TRANSMEM 816 835 POTENTIAL.

FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 148 148 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 859 AA; 97066 MW; 982A9F5A1AD8FA4D CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;

Best Local Similarity 77.8%; Pred. No. 1.3e-02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSVVIVGRI 9

Db 801 GLVIVGRI 809

RESULT 14

ENV\_EIAV9

ID ENV\_EIAV9 STANDARD; PRT; 859 AA.

AC P11306;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein

GP90; Coat protein GP45].

GN ENV.

OS Equine infectious anemia virus (Clone 1369) (EIAV).

OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11670;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87236196; PubMed=3035786;

RA Kawakami T., Sherman L., Dahlberg J., Gazit A., Yaniv A.,

RA Tronick S.R., Aaronson S.A.;

RT "Nucleotide sequence analysis of equine infectious anemia virus

proviral DNA.\*;

RL Virology 158:300-312(1987).

CC [2]

CC REVISIONS TO N-TERMINUS.

CC Tronick S.R.;

RL Submitted (NOV-1987) to the EMBL/GenBank/DBJ databases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M16575; AAB59863.1; -

DR InterPro: IPR001027; GP45\_EIAV.

DR InterPro: IPR001361; GP90\_EIAV.

DR Pfam; PF01045; EIAV\_GP45; 1.

DR Pfam; PF00971; EIAV\_GP90; 1.

KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 859 ENV POLYPROTEIN.

FT CHAIN 23 444 COAT PROTEIN GP90.

FT CHAIN 445 859 COAT PROTEIN GP45.

FT TRANSMEM 75 93 POTENTIAL.

FT TRANSMEM 446 462 POTENTIAL.

FT TRANSMEM 614 636 POTENTIAL.

FT TRANSMEM 787 807 POTENTIAL.

FT TRANSMEM 816 835 POTENTIAL.  
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 859 AA; 97113 MW; 484ED8518CDAF364 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSWIVIGRI 9

I :|:|:|:|

Db 801 GLVIVIGRI 809

## RESULT 15

ENV\_EIAYC

ID ENV\_EIAYC STANDARD; PRT; 859 AA.  
AC P32541;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein  
GP90; Coat protein GP45].  
GN ENV.  
OS Equine infectious anemia virus (clone CL22) (EIAV).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=31675;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92292230; PubMed=1318398;  
RA Perry S.T., Flaherty M.T., Kelley M.J., Clabough D.L., Tronick S.R.,  
Coggins L., Whetter L., Lengel C.R., Fuller F.;  
RT "The surface envelope protein gene region of equine infectious anemia  
virus is not an important determinant of tropism in vitro.";  
RL J. Virol. 66:4085-4097(1992).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M87581; AAA43005.1; -  
DR PIR; C41991; VCLJ22.  
DR InterPro; IPR001027; GP45\_EIAY.  
DR InterPro; IPR001361; GP90\_EIAY.  
DR Pfam; PF01045; EIAV\_GP45\_1.  
DR Pfam; PF00971; EIAV\_GP90\_1.  
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 859 ENV POLYPROTEIN.  
FT CHAIN 23 444 COAT PROTEIN GP90.  
FT CHAIN 445 859 COAT PROTEIN GP45.  
FT TRANSMEM 75 93 POTENTIAL.  
FT TRANSMEM 446 472 POTENTIAL.

FT TRANSMEM 617 636 POTENTIAL.  
FT TRANSMEM 787 807 POTENTIAL.  
FT CARBOHYD 816 835 POTENTIAL.  
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 483 483 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 490 490 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 859 AA; 97140 MW; 23E020E80DF334FA CRC64;

Query Match 69.4%; Score 34; DB 1; Length 859;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSWIVIGRI 9

I :|:|:|:|

Db 801 GLVIVIGRI 809

Search completed: August 30, 2003, 19:13:51

Job time : 1.54479 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:00:22 ; Search time 2.09905 seconds

(without alignments)  
1352.314 Million cell updates/sec

Title: US-09-965-594-26

Perfect score: 49

Sequence: 1 GSVVIGRVL 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rudent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	3010	12 Q9QP06	Q9qp06 hepatitis c
2	49	100.0	3010	12 Q9QP61	Q9qp61 hepatitis c
3	49	100.0	3010	12 Q81760	Q81760 hepatitis c
4	49	100.0	3010	12 Q9DTE5	Q9dte5 hepatitis c
5	49	100.0	4040	12 Q9IFH8	Q9ifh8 mucosal dis
6	48	98.0	88	12 Q39914	Q39914 hepatitis c
7	48	98.0	89	12 Q39895	Q39895 hepatitis c
8	48	98.0	102	12 Q9PXP5	Q9pxp5 hepatitis c
9	48	98.0	138	12 Q68209	Q68209 hepatitis c
10	48	98.0	138	12 Q68235	Q68235 hepatitis c
11	48	98.0	138	12 Q68218	Q68218 hepatitis c
12	48	98.0	138	12 Q68244	Q68244 hepatitis c
13	48	98.0	138	12 Q68206	Q68206 hepatitis c
14	48	98.0	138	12 Q68242	Q68242 hepatitis c
15	48	98.0	138	12 Q68240	Q68240 hepatitis c
16	48	98.0	138	12 Q68213	Q68213 hepatitis c

17	48	98.0	138	12 Q68210	Q68210 hepatitis c
18	48	98.0	138	12 Q68227	Q68227 hepatitis c
19	48	98.0	138	12 Q68229	Q68229 hepatitis c
20	48	98.0	138	12 Q68207	Q68207 hepatitis c
21	48	98.0	138	12 Q68216	Q68216 hepatitis c
22	48	98.0	138	12 Q68228	Q68228 hepatitis c
23	48	98.0	138	12 Q68221	Q68221 hepatitis c
24	48	98.0	138	12 Q68205	Q68205 hepatitis c
25	48	98.0	138	12 Q68215	Q68215 hepatitis c
26	48	98.0	138	12 Q68232	Q68232 hepatitis c
27	48	98.0	138	12 Q68208	Q68208 hepatitis c
28	48	98.0	138	12 Q68231	Q68231 hepatitis c
29	48	98.0	138	12 Q68211	Q68211 hepatitis c
30	48	98.0	138	12 Q68238	Q68238 hepatitis c
31	48	98.0	138	12 Q68237	Q68237 hepatitis c
32	48	98.0	138	12 Q68217	Q68217 hepatitis c
33	48	98.0	138	12 Q68230	Q68230 hepatitis c
34	48	98.0	172	12 Q81579	Q81579 hepatitis c
35	48	98.0	172	12 Q81575	Q81575 hepatitis c
36	48	98.0	172	12 Q81577	Q81577 hepatitis c
37	48	98.0	172	12 Q81582	Q81582 hepatitis c
38	48	98.0	172	12 Q81584	Q81584 hepatitis c
39	48	98.0	172	12 Q81574	Q81574 hepatitis c
40	48	98.0	172	12 Q81578	Q81578 hepatitis c
41	48	98.0	172	12 Q81583	Q81583 hepatitis c
42	48	98.0	172	12 Q81581	Q81581 hepatitis c
43	48	98.0	271	12 Q81573	Q81573 hepatitis c
44	48	98.0	425	12 Q68344	Q68344 hepatitis c
45	48	98.0	1186	12 Q81755	Q81755 hepatitis c

#### ALIGNMENTS

RESULT 1

Q9QP06 PRELIMINARY; PRT: 3010 AA.

ID Q9QP06; AC Q9QP06; DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Genome polypeptide.

OS Hepatitis C virus type 1b.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=31647;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RB;

RA Bartschlagler R.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=RB;

RX MEDLINE=99370154; PubMed=10438800;

RA Koch J.O., Bartschlagler R.;

RT "Modulation of hepatitis C virus NS5A hyperphosphorylation by nonstructural proteins NS3, NS4A, and NS4B.";

RL J. Virol. 73:7138-7146(1999).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; AJ238800; CAB53095.1; -.

DR HSSP; P26663; INS3.

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV\_capsid.

DR InterPro: IPR002521; HCV\_core.

DR InterPro: IPR002519; HCV\_env.

DR InterPro: IPR002531; HCV\_NS1.

DR InterPro: IPR002518; HCV\_NS2.

DR InterPro: IPR004109; HCV\_NS3.

DR InterPro: IPR000745; HCV\_NS4a.

DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RDRP.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00998; Viral\_RDRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART: SM00487; DEXDC; 1.  
DR PROSITE: PS50507; RDRP\_POSITIVE; 1.  
DR PROSITE: PS50521; RDRP\_VIRAL; 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
FT CHAIN 1 191  
FT CHAIN 192 383  
FT CHAIN 384 746  
FT CHAIN 747 809  
FT CHAIN 810 1026  
FT CHAIN 1027 1657  
FT CHAIN 1658 1711  
FT CHAIN 1712 1972  
FT CHAIN 1973 2418  
FT CHAIN 2419 3010  
SQ SEQUENCE 3010 AA; 336999 MW; A570B980DD64634 CRC64;

Query Match 100.0%; Score 49; DB 12; Length 3010;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVIVIGRVL 11  
DB 1678 GSVIVIGRVL 1688  
RESULT 2  
Q9Q61 PRELIMINARY: PRT; 3010 AA.  
AC Q9Q61: 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN SEQUENCE FROM N.A.  
RP STRAIN=274933RU;  
RA Mokhov V.V., Samokhvalov E.I., Novikov D.V., Shatalov A.G.,  
RA Prilipov A.G.;  
RT "Molecular cloning HCV Russian isolate 1b from the serum of patient  
with acute hepatitis.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND MRNA (BY SIMILARITY).  
CC EMBL: AF176573; AAD50312.1; -.  
DR HSP: P26663; INS3.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.

DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RDRP.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV\_capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV\_env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00998; Viral\_RDRP; 1.  
DR ProDom: PD186062; HCV\_NS1; 1.  
DR SMART: SM00487; DEXDC; 1.  
DR PROSITE: PS50507; RDRP\_POSITIVE; 1.  
DR PROSITE: PS50521; RDRP\_VIRAL; 1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
Hydrolase; Nonstructural protein; Polyprotein;  
RNA-directed RNA polymerase; Transferase; Transmembrane.  
SQ SEQUENCE 3010 AA; 327068 MW; 9105F69493DD5BBA CRC64;

Query Match 100.0%; Score 49; DB 12; Length 3010;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVIVIGRVL 11  
DB 1678 GSVIVIGRVL 1688

RESULT 3  
Q81760 PRELIMINARY: PRT; 3010 AA.  
AC Q81760: 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DE Genome polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN SEQUENCE FROM N.A.  
RP STRAIN=HC-C2;  
RX MEDLINE=93359897; PubMed=8394876;  
RA Wang Y., Okamoto H., Tsuda F., Nagayama K., Tao Q.M., Mishiro S.;  
RT "Prevalence, Genotypes, and an isolate(HC-C2) of Hepatitis C Virus in  
Chinese Patients With Liver Disease.";  
RL J. Med. Virol. 40:254-260(1993).  
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND MRNA (BY SIMILARITY).  
CC EMBL: D10934; BAA01728.1; -.  
DR HSP: P26663; INS3.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR002522; HCV capsid.  
DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV\_NS1.  
DR InterPro: IPR002518; HCV\_NS2.  
DR InterPro: IPR004109; HCV\_NS3.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.

DR	InterPro: IPR002868; HCV_NS5a.
DR	InterPro: IPR002166; HCV_RdRP.
DR	InterPro: IPR007095; RNA_pol_DS_PS.
DR	InterPro: IPR007094; RNA_pol_PSVir.
DR	Pfam: PF01543; HCV_capsid; 1.
DR	Pfam: PF01542; HCV_core; 1.
DR	Pfam: PF01539; HCV_env; 1.
DR	Pfam: PF01560; HCV_NS1; 1.
DR	Pfam: PF01538; HCV_NS2; 1.
DR	Pfam: PF02907; HCV_NS3; 1.
DR	Pfam: PF01006; HCV_NS4a; 1.
DR	Pfam: PF01001; HCV_NS4b; 1.
DR	Pfam: PF01506; HCV_NS5a; 1.
DR	Pfam: PF00998; Viral_RdRP; 1.
DR	ProDom: PD186062; HCV_NS1; 1.
DR	SMART: SM00487; DEXDC; 1.
DR	PROSITE: PS05057; RDRP_POSITIVE; 1.
DR	PROSITE: PS05021; RDRP_VIRAL; 1.
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW	Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SEQ	SEQUENCE 3010 AA; 326855 MW; EAYD306A4BAZE224 CRC64;

  

Query Match	100.0%;	Score 49;	DB 12;	Length 3010;
Best Local Similarity	100.0%;	Pred. NO. 6.6;		
Matches 11;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

  

QY	1	GSWVIVGRVL 11
DB	1678	GSWVIVGRVL 1688

  

RESULT 4	
Q90TES	PRELIMINARY: PRT; 3010 AA.
ID	Q9DTE5
AC	Q9DTE5
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Genome polyprotein.
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepacivirus.
OX	NCBI_TaxID=11103;
RN	{1}
RP	SEQUENCE FROM N.A.
RC	STRAIN=HCV145;
RA	Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA	Mishihara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA	Mishiro S.;
RT	"Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT	with hepatocellular carcinoma: the progression score' revisited."
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC	LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC	PROTEIN C AND MRNA (BY SIMILARITY).
DR	ENBL; AB049092; BAB18805.1;
DR	HSP; P27958; 1HEI.
DR	InterPro: IPR000345; CytC_heme_bind.
DR	InterPro: IPR001410; DEAD.
DR	InterPro: IPR002522; HCV_capsid.
DR	InterPro: IPR002521; HCV_core.
DR	InterPro: IPR002519; HCV_env.
DR	InterPro: IPR002531; HCV_NS1.
DR	InterPro: IPR002518; HCV_NS2.
DR	InterPro: IPR004109; HCV_NS3.
DR	InterPro: IPR000745; HCV_NS4a.
DR	InterPro: IPR001490; HCV_NS4b.
DR	InterPro: IPR002868; HCV_NS5a.
DR	InterPro: IPR002166; HCV_RdRP.
DR	InterPro: IPR001650; Helicase_C.
DR	InterPro: IPR007095; RNA_pol_DS_PS.
DR	InterPro: IPR007094; RNA_pol_PSVir.

```

DR PROSITE; P50507; RDRP_POSITIVE; 1.
DR PROSITE; P50521; RDRP_VIRAL; 1.
DR PROSITE; P50531; RNASE_I2_2; 1.
DR ATP-binding; Helicase; Hydrolase; Nonstructural protein; Polyprotein;
RNA-directed RNA polymerase; Transferase.
SQ SEQUENCE 4040 AA; 453073 MW; ADE87791D055B9DC CRC64;

Query Match 100.0%; Score 49; DB 12; Length 4040;
Best Local Similarity 100.0%; Pred. No. 8.8; Mismatches 0; Gaps 0;
Matches 11; Conservative 0; Indels 0;

Qy 1 GSVVIVGRIVL 11
   |||||
Db 10 GSVVIVGRIVL 20

RESULT 6
O39914 PRELIMINARY; PRT; 88 AA.
AC O39914;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FO1;
RX MEDLINE=98032593; PubMed=9365889;
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 53:237-244(1997).
DR EMBL; AF007519; AAB62970.2; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 9750 MW; BF7B5198B317B6E0 CRC64;

Query Match 98.0%; Score 48; DB 12; Length 88;
Best Local Similarity 90.9%; Pred. No. 0.3; Mismatches 1; Gaps 0;
Matches 10; Conservative 1; Indels 0;

Qy 1 GSVVIVGRIVL 11
   |||||
Db 6 GSVVIVGRIVL 16

RESULT 7
O39895 PRELIMINARY; PRT; 89 AA.
AC O39895;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of Hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";

```

```

RL J. Med. Virol. 0:0-0(1997).
DR EMBL; AF007500; AAB62951.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA; 9748 MW; 581BB8C8A3EA8B5C CRC64;

Query Match 98.0%; Score 48; DB 12; Length 89;
Best Local Similarity 90.9%; Pred. No. 0.3; Mismatches 1; Gaps 0;
Matches 10; Conservative 1; Indels 0;

Qy 1 GSVVIVGRIVL 11
   |||||
Db 8 GSVVIVGRIVL 18

RESULT 8
O9PXP5 PRELIMINARY; PRT; 102 AA.
AC O9PXP5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Non-structural protein NS4-GROUP II HCV-specific antigen C14-1
DE (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94245087; PubMed=7514558;
RA Tanaka T., Tsukiyama-Kohara K., Yanaguchi K., Yagi S., Tanaka S.,
RA Hasegawa A., Ohta Y., Hattori N., Kohara M.;
RT "Significance of specific antibody assay for genotyping of hepatitis C
RT virus.";
RL Hepatology 19:1347-1353(1994).
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 102
FT NON_TER 102
SQ SEQUENCE 102 AA; 11419 MW; 08124C19CF367F06 CRC64;

Query Match 98.0%; Score 48; DB 12; Length 102;
Best Local Similarity 90.9%; Pred. No. 0.35; Mismatches 1; Gaps 0;
Matches 10; Conservative 1; Indels 0;

Qy 1 GSVVIVGRIVL 11
   |||||
Db 20 GSVVIVGRIVL 30

RESULT 9
O68209 PRELIMINARY; PRT; 138 AA.
AC O68209;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1b;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
RT Asia.";
RL J. Gen. Virol. 76:211-215(1995).

```

DR EMBL; U14245; AAC53934.1; -  
DR HSSP; P27958; IHEI.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138  
FT NON\_TER 138  
SQ SEQUENCE 138 AA; 15149 MW; DBAE62A0FE9E2D57 CRC64;  
Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSVVIVGRIVL 11  
Db 52 GSVVIVGRIL 62  
RESULT 10  
Q68235 PRELIMINARY; PRT; 138 AA.  
AC Q68235;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1b;  
RX MEDLINE=95146953; PubMed=7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;  
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia."  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL; U14271; AAC53960.1; -  
DR HSSP; P27958; IHEI.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138  
FT NON\_TER 138  
SQ SEQUENCE 138 AA; 15104 MW; 585DC5A627D0F3E3 CRC64;  
Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSVVIVGRIVL 11  
Db 52 GSVVIVGRIL 62  
RESULT 11  
Q68218 PRELIMINARY; PRT; 138 AA.  
AC Q68218;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1b;  
RX MEDLINE=95146953; PubMed=7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;

RT "Prevalence of hepatitis C virus sequence variants in South-East Asia."  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL; U14254; AAC53943.1; -  
DR HSSP; P27958; IHEI.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138  
FT NON\_TER 138  
SQ SEQUENCE 138 AA; 15189 MW; DB78E92DDC67040F CRC64;  
Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSVVIVGRIVL 11  
Db 52 GSVVIVGRIL 62  
RESULT 12  
Q68244 PRELIMINARY; PRT; 138 AA.  
AC Q68244;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1b;  
RX MEDLINE=95146953; PubMed=7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;  
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia."  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL; U14280; AAC53969.1; -  
DR HSSP; P27958; IHEI.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138  
FT NON\_TER 138  
SQ SEQUENCE 138 AA; 15118 MW; B7F7EB2733770408 CRC64;  
Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSVVIVGRIVL 11  
Db 52 GSVVIVGRIL 62  
RESULT 13  
Q68206 PRELIMINARY; PRT; 138 AA.  
AC Q68206;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.



RC STRAIN=1b;  
RX MEDLINE-95146953; PubMed-7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;  
RT "Prevalence of hepatitis C virus sequence variants in South-East  
RT Asia.";  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL: U14276; AAC53965.1; -.  
DR HSSP: P27958; 1HEI.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138 138  
SQ SEQUENCE 138 AA; 15117 MW; 5FBC51A1B74DE13E CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 1;

QY 1 GSVWIVGRIVL 11  
Db 52 GSVWIVGRILL 62  
|||||||:|

## RESULT 14

Q68242  
ID Q68242 PRELIMINARY; PRT; 138 AA.  
AC Q68242;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1b;  
RX MEDLINE-95146953; PubMed-7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;  
RT "Prevalence of hepatitis C virus sequence variants in South-East  
RT Asia.";  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL: U14276; AAC53967.1; -.  
DR HSSP: P27958; 1HEI.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138 138  
SQ SEQUENCE 138 AA; 15281 MW; CDSB5B3834C6070D CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 1;

QY 1 GSVWIVGRIVL 11  
Db 52 GSVWIVGRILL 62  
|||||||:|

## RESULT 15

Q68240  
ID Q68240 PRELIMINARY; PRT; 138 AA.  
AC Q68240;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Nonstructural protein (Fragment).  
GN NS4.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.

OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=1b;  
RX MEDLINE-95146953; PubMed-7844535;  
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;  
RT "Prevalence of hepatitis C virus sequence variants in South-East  
RT Asia.";  
RL J. Gen. Virol. 76:211-215(1995).  
DR EMBL: U14276; AAC53965.1; -.  
DR HSSP: P27958; 1HEI.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
FT NON\_TER 1  
FT NON\_TER 138 138  
SQ SEQUENCE 138 AA; 15115 MW; 6A042677B354CA7A CRC64;

Query Match 98.0%; Score 48; DB 12; Length 138;  
Best Local Similarity 90.9%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 1;

QY 1 GSVWIVGRIVL 11  
Db 52 GSVWIVGRILL 62  
|||||||:|

Search completed: August 30, 2003, 19:18:23  
Job time : 4.09905 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:18:33 ; Search time 142.976 Seconds  
(without alignments)  
3147.423 Million cell updates/sec

Title: US-09-965-594-26  
Perfect score: 49  
Sequence: 1 GSWIVGRIVL 11

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09965594/runat\_29082003.151919.28310/app\_query.fasta.1.2872  
-DB=GenEmbl -QFMT=fastap -SURFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.coi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORX=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGN=1.14686 -runat\_29082003.151919.28310 -NCPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgtg\_other:\*  
33: em\_hgtg\_mus:\*  
34: em\_hgtg\_pln:\*  
35: em\_hgtg\_rod:\*  
36: em\_hgtg\_mam:\*  
37: em\_hgtg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_mus:\*  
41: em\_hgtgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	618	6	E06198
2	49	100.0	618	6	E06394
3	49	100.0	7475	6	A91965
4	49	100.0	7475	6	AR031992
5	49	100.0	7475	6	AR207294
6	49	100.0	9033	14	HCJ238800
7	49	100.0	9344	14	AB049092
8	49	100.0	9400	14	HPCGGENOM
9	49	100.0	9456	14	HPCRNA
10	49	100.0	9600	14	AF176573
11	49	100.0	12734	6	AR179057
12	49	100.0	12734	14	AF268278
13	48	98.0	75	6	AR145217
14	48	98.0	75	6	AR145221
15	48	98.0	78	6	AR145197
16	48	98.0	78	6	AR145213
17	48	98.0	96	6	AR145247
18	48	98.0	96	6	AR145249
19	48	98.0	161	6	AX481515
20	48	98.0	189	6	AR037527
21	48	98.0	189	6	E09290
22	48	98.0	266	14	AF007519
23	48	98.0	267	6	AR037528
24	48	98.0	267	6	E09291
25	48	98.0	267	14	AF007500
26	48	98.0	279	6	E11063
27	48	98.0	321	6	E11062
28	48	98.0	372	6	E03389
29	48	98.0	414	14	HCUI4241
30	48	98.0	414	14	HCUI4242
31	48	98.0	414	14	HCUI4243
32	48	98.0	414	14	HCUI4244
33	48	98.0	414	14	HCUI4245
34	48	98.0	414	14	HCUI4246
35	48	98.0	414	14	HCUI4247
36	48	98.0	414	14	HCUI4249
37	48	98.0	414	14	HCUI4251
38	48	98.0	414	14	HCUI4252
39	48	98.0	414	14	HCUI4253
40	48	98.0	414	14	HCUI4254
41	48	98.0	414	14	HCUI4257
42	48	98.0	414	14	HCUI4263
43	48	98.0	414	14	HCUI4264
44	48	98.0	414	14	HCUI4265
45	48	98.0	414	14	HCUI4266

ALIGNMENTS

E06198  
LOCUS E06198 618 bp RNA linear PAT 29-SEP-1997  
DEFINITION cDNA encoding genes derived from hepatitis C virus.  
ACCESSION E06198  
VERSION E06198.1 GI:2174385  
KEYWORDS JP 1994000085-A/38.  
SOURCE Hepatitis C virus  
ORGANISM Hepatitis C virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
REFERENCE 1 (bases 1 to 618)  
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.  
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION  
JOURNAL Patent: JP 1994000085-A 38 11-JAN-1994;  
COMMENT MITSUBISHI KASEI CORP  
OS (hepatitis C virus)  
PN JP 1994000085-A/38  
PD 11-JAN-1994  
PF 11-JUN-1992 JP 1992194497  
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957  
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,  
PI TERANISHI YUTAKA, HAYASHI NORIO  
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC C12N1/21,C12N5/10,  
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/29, PC (C12N1/21), (C12N5/10,C12R1:91), (C12P21/02,C12R1:19), (C12P21/02, C12R1:91), (C12N5/10,C12R1:91), (C12P21/02,C12R1:19), (C12P21/02, C07K99:00;  
PC C07K99:00;  
CC strandedness: Double;  
CC topology: Linear;  
CC anti-sense: NO;  
CC \*source: clone-N13-1;  
FH Key Location/Qualifiers  
FH mat\_peptide 9..608  
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis patient'.  
FT type C  
FT Location/Qualifiers  
FT 1..618  
FT /organism='Hepatitis C virus'  
FT /mol\_type='genomic RNA'  
FT /db\_xref='taxon:11103'  
BASE COUNT 135 a 184 c 179 g 120 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.238 Length: 618  
Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-26 (1-11) x E06198 (1-618)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11  
|||||  
Db 351 GCCAGCGTGCATTGTGGCAGGATCGTCTTG 383

RESULT 2  
E06198  
LOCUS E06198 618 bp RNA linear PAT 29-SEP-1997  
DEFINITION cDNA encoding genes derived from hepatitis C virus.  
ACCESSION E06198

E06394.1 GI:2174581  
KEYWORDS JP 1994000086-A/38.  
SOURCE Hepatitis C virus  
ORGANISM Hepatitis C virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  
REFERENCE 1 (bases 1 to 618)  
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.  
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION  
JOURNAL Patent: JP 1994000086-A 38 11-JAN-1994;  
COMMENT MITSUBISHI KASEI CORP  
OS (hepatitis C virus)  
PN JP 1994000086-A/38  
PD 11-JAN-1994  
PF 07-OCT-1992 JP 1992293734  
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR 20-APR-1992 JP 92P 99957  
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI MURAKAMI TOMOKO,  
PI TERANISHI YUTAKA, HAYASHI NORIO  
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC C12N15/11,  
PC C12N15/85,C12P21/02//A61K39/29, (C12P21/02,C12R1:91),C07K99:00;  
CC strandedness: Double;  
CC topology: Linear;  
CC anti-sense: NO;  
CC \*source: clone-N13-1;  
FH Key Location/Qualifiers  
FH mat\_peptide 9..608  
FT /product='the peptides reacting specifically and immunochemically with the serum of hepatitis patient'.  
FT type C  
FT Location/Qualifiers  
FT 1..618  
FT /organism='Hepatitis C virus'  
FT /mol\_type='genomic RNA'  
FT /db\_xref='taxon:11103'  
BASE COUNT 135 a 184 c 179 g 120 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.238 Length: 618  
Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-965-594-26 (1-11) x E06394 (1-618)

Qy 1 GlySerValValIleValGlyArgIleValLeu 11  
|||||  
Db 351 GCCAGCGTGCATTGTGGCAGGATCGTCTTG 383

RESULT 3  
A91965  
LOCUS A91965 7475 bp DNA circular PAT 22-JAN-2000  
DEFINITION Sequence 1 from Patent WO9822496.  
ACCESSION A91965  
VERSION A91965.1 GI:6740811  
KEYWORDS unidentified  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 7475)  
AUTHORS Attwood,M.R. and Hurst,D.N.  
TITLE ANTIVIRAL PEPTIDE DERIVATIVES  
JOURNAL Patent: WO 9822496-A 1 28-MAY-1998;

```

FEATURES             Location/Qualifiers
  source              1..7475
                    /organism="unidentified"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32644"
BASE COUNT          1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:             49.00      Matches:     11
Score:                 100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:       0
Query Match:           100.00%     Gaps:         0
DB:
US-09-965-594-26 (1-11) x AR031992 (1-7475)

Qy      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATGTGGCAGGATCGTCTTG 3483

RESULT 4
LOCUS      AR031992              7475 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5866684.
ACCESSION  AR031992
VERSION     AR031992.1 GI:5946281
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 7475)
AUTHORS   Attwood,M.Richard., Hurst,D.Nigel., Jones,P.Stephen.,
           Kay,P.Brittain., Raynham,T.Michael. and Wilson,F.Xavier.
TITLE     Peptidyl inhibitors of viral proteases
JOURNAL   Patent: US 5866684-A 1 02-FEB-1999;
FEATURES   Location/Qualifiers
  source    1..7475
            /organism="unknown"
BASE COUNT  1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:             49.00      Matches:     11
Score:                 100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:       0
Query Match:           100.00%     Gaps:         0
DB:
US-09-965-594-26 (1-11) x AR031992 (1-7475)

Qy      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATGTGGCAGGATCGTCTTG 3483

RESULT 5
LOCUS      AR207294              7475 bp      DNA      linear      PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6372883.
ACCESSION  AR207294
VERSION     AR207294.1 GI:21506162
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 7475)
AUTHORS   Attwood,M.Richard., Hurst,D.Nigel., Jones,P.Stephen.,
           Kay,P.Brittain., Raynham,T.Michael. and Wilson,F.Xavier.
TITLE     Antiviral medicaments
JOURNAL   Patent: US 6372883-A 1 16-APR-2002;
FEATURES   Location/Qualifiers
  source    1..7475
            /organism="unknown"
BASE COUNT  1853 a 1916 c 2039 g 1667 t
ORIGIN
Alignment Scores:      3.53      Length:      7475
Pred. No.:             49.00      Matches:     11
Score:                 100.00%     Conservative: 0
Percent Similarity:    100.00%     Mismatches:  0
Best Local Similarity: 100.00%     Indels:       0
Query Match:           100.00%     Gaps:         0
DB:
US-09-965-594-26 (1-11) x AR031992 (1-7475)

Qy      1 GlySerValValIleValGlyArgIleValLeu 11
Db      3451 GGCAGCGTGTTCATGTGGCAGGATCGTCTTG 3483

RESULT 6
LOCUS      HCJ238800              9033 bp      RNA      linear      VRL 18-AUG-1999
DEFINITION Hepatitis C virus type 1b complete genome, isolate NCI.
ACCESSION  AJ238800
VERSION     AJ238800.1 GI:5748510
KEYWORDS   complete genome; core protein; glycoprotein E1; glycoprotein E2;
           NS2 proteinase; NS3 proteinase/helicase; NS3/4A proteinase
           cofactor; NS4b protein; NS5A phosphoprotein; NS5B RNA dependant RNA
           polymerase; p7 peptide; polyprotein.
SOURCE     Hepatitis C virus type 1b
ORGANISM   Hepatitis C virus type 1b
           Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
           Hepacivirus.
REFERENCE  1
AUTHORS   Koch,J.O. and Bartenschlager,R.
TITLE     Modulation of hepatitis C virus NS5A hyperphosphorylation by
           nonstructural proteins NS3, NS4A, and NS4B
JOURNAL   J. Virol. 73 (9), 7138-7146 (1999)
MEDLINE   99370154
PUBMED    10438800
REFERENCE  2 (bases 1 to 9033)
AUTHORS   Bartenschlager,R.
TITLE     Direct Submission
JOURNAL
FEATURES   Location/Qualifiers
  source    1..9033
            /organism="Hepatitis C virus type 1b"
            /viroion
            /mol_type="genomic RNA"
            /strain="RB"
            /isolate="NCI"
            /db_xref="taxon:31647"
            /country="Germany"
            /codon_start=1
            /product="polyprotein"
            /protein_id="CA53095.1"
            /db_xref="GI:5748511"
            /translation="MSTNPKPQKTKRNTNRRPDQVKFPGGGQIVGGVYLLPRRGPRL
            GYVTRKTSRSGPRGRQPIPKARQEGRAWAQGPWPLYGNEGGLGWAGWLLSPRG
            SRPSGPTDPRRSRLGVIDTLCGFADLMGLPLVAGPLGGAARALAHGVRVLED
            GYNATGNLPGCSFSIFLLALLSCLTIPASAYEVNRVGVYHVYVNDCSNASIVYEAD
            NIMHTPGVCYCVRENNSSRCWALPTLAARNASVPTTIRHVDLLVGAAALCSAMY
            VQLDGSFVLAQLFTFSPPRHTVODKCSYIPGHVTHGRMAWMMNMWSPALVW
            SGLLRIPQAVVDNVAHGWVLAGLAYISVGNWAKVLIWLLIFAGVDGTTVTGTM
            AKNTIGTSLFSPGSSQKIQLVNTNGSHINRTALNCNDSLNTGLAALFYVVKNS
            GCPERMASGSPIDAFAGWGPITYNESHSDDORPYCMHYAPRPGIVPAOVCVPYC
            FTPSPVVGTTDRFGVPTYTWSGENETDVLNNTRPQGNWFCCTWMSGTGFTKCGG
            PCNIGGSKLTCTPDCFRRKPEATYTKCGSGPWLIPRCLVHPYPLRWHPYCTVNF
            TIFKVRMYGGVEHRLAEACNWTGRERCNLEDRDRSELSPILLSTTEMQVLPFCSTTL

```





Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x HPCRNA (1-9456)

QY 1 GlySerValValIleValGlyArgIleValLeu 11  
|||||  
DB 5373 GCGCATGTGGTCATTTGGCGAGGATCGTCTTG 5405

RESULT 10  
LOCUS AF176573 9600 bp RNA linear VRL 18-AUG-1999  
DEFINITION Hepatitis C virus polyprotein precursor, gene, complete cds.  
ACCESSION AF176573  
VERSION AF176573.1 GI:5738246  
KEYWORDS  
SOURCE  
ORGANISM  
Hepatitis C virus  
Hepatitis C virus  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
Hepadnaviruses.

REFERENCE 1 (bases 1 to 9600)  
AUTHORS Mokhonov, V.V., Samokhvalov, E.I., Novikov, D.V., Shatalov, A.G. and Prilipov, A.G.  
TITLE Molecular cloning HCV Russian isolate lb from the serum of patient with acute hepatitis  
JOURNAL Unpublished  
AUTHORS  
REFERENCE 2 (bases 1 to 9600)  
AUTHORS Mokhonov, V.V., Samokhvalov, E.I., Novikov, D.V., Shatalov, A.G. and Prilipov, A.G.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-1999) Molecular Genetics, Ivanovsky Virology Institute, Gamaleya Str., 16, Moscow 123098, Russia

FEATURES  
source  
1..9600  
/organism="Hepatitis C virus"  
/mol\_type="genomic RNA"  
/strain="274933R0"  
/isolate="1b"  
/db\_xref="taxon:11103"  
/country="Russia"  
/note="Isolated from acute hepatitis patient serum"  
1..341  
342..9374  
/note="cleaved into C, E1, E2, p7, NS2, NS3, NS4a, NS4b, NS5a, and NS5b proteins"  
/codon\_start=1  
/product="polyprotein precursor"  
/protein\_id="AA050312.1"  
/db\_xref="GI:5738247"

5'UTR  
CDS

1..9456  
/organism="Hepatitis C virus"  
/mol\_type="genomic RNA"  
/isolate="HC-C2"  
/db\_xref="taxon:11103"  
1..341  
342..9374  
/note="C, E, NS1/E2, NS3, NS4 and NS5"  
/codon\_start=1  
/product="polyprotein precursor"  
/protein\_id="BAA01728.1"  
/db\_xref="GI:471117"  
/translation="MSTNPKPQKTKNTNRPDQVKFPGCGOIVGGVYLLPRGPRL  
GYRATKTSERSQRRQPIKARREGRANAAQPGYPMPLYGNEGHWAGHLLSPRG  
SRSPGNDPRRRRNKGKVIDTLTCGFADLMGYIPLVGAFLGGAAALAHGVYAEAD  
GVNATGLPGCSFSIFLLALLSCLTIPASAEYRNASGVYHVNDCNSIVYEAAD  
MIMNPGCVPCRRNRSQWVALPTLAARNASVPTTIRRHVDLLVGAALCSANY  
VMDGCGVFLVQLFTSPRRHETVQDCNSIYPGHITGRMAWMMNNKSPITLAVY  
SOLLRIPOAVMDVAGAHGVLAGLAVYSMVGNKAKVLI VLLLFAGVGGNTYVYTGAA  
ARGGKISLFSRGPQKIQLVNTYNSWHINRTALNCNDSFNTGFLAAALFYARFNSS  
CQPERMASCSISIDFDQWGPITYYQDSDQRPYCHYIPRPGCIIPVASEVCGPYC  
FTPSPVVGVTDRGLGVTYNGENEDVLLNNTPRPGNWFCTNNTHYFTKTCGG  
PPCINAGNNTLTCTDCFRKHEATYTKCGSPMLTPCLDYIPLRMHTCYTNE  
TIFYRVYGVGVEHRLDAACNWTGERCALDEDRSELSPLLLSTETWELPCSFSTL  
PALSTGTLHNRHVDVQYLYIGSAGVSAIKWEYVLLFLLADARVCACMLMLL  
LQAALPRAYAMAREMASCGAVFVGLALLTISPYVKFLARLIMWLOYFITRAEHL  
LQWLPPLNVGGRDAIILTCAAHPELIIDITIKLLAIFGPIIMVQAGITKMPYFVR  
ACGLRACMKVRKGVGYOMAKLAAUTGTYIVDHLTPLRDMAHTGRLDAVAE  
PVFSDEMTEK1ITWAGDTAAGDITILGPVSARKEIILGPADSLGGQWRLLAPIT  
YLAQTPKGLGCIITSLTRDKNOVEGVVSTATQSFATCVNGVCTVYHYGARGK  
TISQTPKGIOTMYITNVQDILVGPAPGARSALTPTCGSSDLYLVTRYADIVPRRG  
MRSPLTSPRVSYLKGSGGPLLCPGSHAAGIPRAAVCTRGVAKAVFVNPVSMETT  
DSRPFVDSPPAVPOTFOVAHLRAPTGSKTKVPAAYAAOGYKVLVNPVSAATL  
GFGAYMSAHGIDPNTRGVYTIITGAPITYSTYKFLADGGSGGNYDIIICDECHS  
ADPITILGGHLLIFQAEATAGRLVLAIRTPPGSYVTPHNSIEVALSYITGEIPFTYK  
TSLTIGKGRHLLIFQAEATAGRLVLAIRTPPGSYVTPHNSIEVALSYITGEIPFTYK  
TDALMTFTGDSFVIDNCTVDFSLDPTFIETTPQDASRSQRRGTGRG  
RGRYFTPTGPERSGMFDSCVLCYDAGCAWYELTPAETSRLRAYLNTLPVQCQ  
DHLEPWSVFTGLTHIDAHFSLQTKQAGDNFPLVAVOATVCARAOAPPSPSWQMMKC  
LIRLQTLRGHTPLLYRLGAVQNEVLTHTPTIKYIMACMSADLEWVTTSTWLYGVYLA  
ALAYCLTGSWIVGRLVSRPAVTPDREYLYQEDMEEGCSHLYIEQGMQALAE  
QFKQALGLQIATKQAEAPAVPESKRWALETFWAKHMNFISGVQYLAGLSTLGN  
PAIASLMAFTASVTSQTSTLFTNIIILGGWAAOLAPPASAAPVAGAGTAAAGVSI  
GLGKVLVDILAGYGAGVAGALVAFKIMSGETPSAEDLVNLLPAILSPGALVGVCAA  
IORRHVGEAGVOMNRLIAFASGRTPSVHYVPESDAAARVOTLSLITQLIK  
RLHONDEDCSTPCSGSDIOWICSVLTDQLQSKLLPLRPGVPFFSCORGYK  
GWMRGDLMOTTCGQAOTGCHKVNGSMRIYVGPKTSNTWHGTFPIINAYTFTGPTSP  
APNTSRALRWRAAEYVEVTVGDFHVTGNTDNKRCQVAPAEFEFTVDGVRLHR  
YAPACKPVIREVDFQVLGKQYVPGSOLPCEPDPVAVLTSMITDPSHITAEAKRRL  
ARGSPSLAASSSOLSAQPSKATCTTHSDPADLIEANLIMQEMGNITRVESN  
KVLVDSEPLRAEEDREVSAAEILRKTRTFPPMPVWADYDNPPLLESKKDPT  
VPVHVGCLPPTKADPIPPRRRTYVLTSTVSSALAEATITFGSSGSAVDST  
ATAPDQTSKDESDAESYSSWPPLEGGDPDLDSGWSVTSEASGVVCCSS  
YTWATGLTPCAAEESKLIPNLSNLLRHNNMYATTISASUOKKVTDFDLOVID  
DHYRDLKEMAKASTVKAKLLSIEACKLTPPHSAKSKPCYKADKVRNLSKAVANI  
RSYWKDLLETETPIDTITMAKNEVFCQPEKGRKPARLIVFPDLGVYCEKMALYD  
VYSTLPQAVMGSSYGFQYSPQGRVEFLVNTWKSKMPGSYDFTRCFDSTVENDIRI  
EESIYQCDLAPAEAKQATKSTERYLIGGLTNSKGNQYRRCASTVLTSCGNTL  
TCYLKASACRAAKLQDLMVNGDOLVIESAGTQEDAAASLVFTTEAMTRYASAPG  
DPPOEYDELIITSCSNVSAHADSGKRYVLTDRPTILPARAAWETARHTPVNSWL  
GNIINYPATLNMARLIMTHPFSILLAOELEKALECOTYGACYSIEPLDLPIELRLH  
GLSASFHSYSGPELNVRVASCRLKGLVPPLEYWRHARSYRAKLLSOGGRAATCGKIL  
FNWAVYTKLUTPIPAASRLSDLSGWVAGYGGGDIYHSLSKARPRWFMCLLILSVGV  
GYLLPNR"

9375..9456  
/note="with a T-stretch of 45 bp."  
1909 a 2611 c 2668 g 2068 t

3'UTR  
BASE COUNT 1909 a 2611 c 2668 g 2068 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.55 Length: 9456

MRSPVFTDNSSPPAVPQTFQVAHLHAPTGSCKSTKVPAAYAAQGYKVLVNPNSVAATL,  
GFGAYMSAHGVDPNLRGTITTTGAPITYSTYCKFLADGGCGGAYDIICDEYCH  
TDSITLIGTVGLQAEAGARLVATATPPGVSVPHPNIEGIALSNTGIEPFCYK  
AIPBILKGGHBLIFCHSKKDELAALSLGLNAVAYRGLDVSVPFSGNVVVA  
TALMTGTFDNCVDTCVTFDSLDPTFIETITVPQDAVSQRGRGTGRG  
RRGIYFVTPGERPSGMSDSSVCEYDAGCAMELTPAETSVRLRAYLNTPLGPVCO  
DHLEFWSVFTGLTHIDAHFSLQTKOAGDNFYLVAOATVCAQAQAPSTPWSQMMKC  
LIRLPTLHGTPPLLYRLGAVONETTLTHPTIKYIMACSADLEVTSTVYLVGGVLA  
ALAAVCLTGSVIVGRVLVSKPAIIPDREVLYOEFDEMECEASHPLPIEOGMOLAE  
OFKALGLLOTATKQAEAAAVVESKWRALAEAFKHMWNFISGVOLAGLSLPGN  
PAISLMAFTASITSPITQYTLLENILGGHVAQAQLAPPSSAASAFVAGANGAAGVSI  
GLKVLVDILAGVAGAGALVAFKVMGSEMPSTEDLVNLLPALISPCALVGVVCAA  
ILRRHVGEAGEAGOMMNLIAFASGNHVSPTHYVPESDAAARQTOILSNVITQILLK  
RLHWINDECTSPGMSLDVMDICTVLTDFKTLRSKLLPRLPGVPLFCQCGYK  
GVMRGDMOTTCGAGIAGHNGSMRIVGRPTCSNTWHTGTFPIINATYTGCTPSP  
APNYSRALWRVAEEYVEITRVGDFHYVTGMTDNVCKPCOVPAPEFTEVDGVLHR  
YAPVCKPILHEDVTFVGLNOYLQVLSQLPCEPEPDVAVLTSMLDPSHITAEAKRL  
ANGSPSLASSASQSLKATCTTTRHSDPADLIEANLLWRQEMGNIRVSEEN  
KVWILDSFDPLRAEEDEREVPVPAELKTRKFPSPIMAPDYNPLLESXKDPDY  
VPVVHGGPLPTKVPPIPPRRKRVTTLTSTVSSALAEATKTFGSESSAVDSGT  
AAASPDQSDNGTDSVSESSMPLEGEFDPDLSDGMSSTVSEASESSVCCSMS  
YTWTCALLTCAAEESOLPINALSNLLRHLNLYATTSRASOROKKVTEDRLQVLD  
PHYOVLEKMKAKASTVKALLSVFEACKLTPPHSAKSEFGYKQDVNLSKAVNHI  
HVMKDLLEDETPTIDTITMAKNEVFCQPEKGGKPARLLVFPDLGVRCCKMALVD  
VYSTLPQAVMSAIGQIYSPGQVDFLYNAWKKCPMGFSFIDTCFDSVTESDIRV  
EESIYQCCDLAPARQAIRSLTERLYVGGPLTNSKQNGYRCRASGLVLTSCGNTL  
TCYLKATAACRAARQDCTLLVGGDOLVICESAGTQBDAAASLRVFTAMTRYAPGP  
DLPOEYDELEITSSSVVAHADSGKRVYLTDRPTPLARAWEAARHTPYSNL  
GNIMYAPTLMWMTLMTFHSILLFQBLEKALDCQIYGAYISIEPLDLQPIIOTLRL  
GLSFLSHSYSGEINRVASCLKGLVPLRAWRHARNVRKALLSOGGRAATCGKYL  
FNMAVTKLKLPIPAASQDLSDGFWAGYGGGDIYHLSLRARPMPWMLCLLLFLFVG  
GLYLPNR"

9375..9600

3'UTR BASE COUNT 1917 a 2887 c 2696 g 2100 t  
ORIGIN

Alignment Scores:

Pred. No.: 4.62 Length: 9600  
Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x AF176573 (1-9600)

Qy 1 GlySerValValleValGlyArgileValLeu 11  
Db 5373 GCGAGGTGTCATCTGGCGCAGGATCGTCTTG 5405

RESULT 11

LOCUS AR179057 12734 bp DNA linear PAT 20-APR-2002

DEFINITION Sequence 1 from patent US 6326137.

ACCESSION AR179057

VERSION AR179057.1 GI:20220612

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 12734)

AUTHORS Hong, Z., Lai, V.C.H. and Lau, J.Y.N.

TITLE Hepatitis C virus protease-dependent chimeric pestivirus

JOURNAL Patent: US 6326137-A 1 04-DEC-2001;

FEATURES Location/Qualifiers

source 1..12734

/organism="unknown"

BASE COUNT 4032 a 2604 c 3295 g 2803 t

ORIGIN

Alignment Scores:

Pred. No.: 6.27 Length: 12734  
Score: 49.00 Matches: 11

Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-09-965-594-26 (1-11) x AR179057 (1-12734)  
Qy 1 GlySerValValleValGlyArgileValLeu 11  
Db 413 GGTAGTGTCTTATTGTTGGTAGAATGTTTAA 445

RESULT 12

LOCUS AF268278

DEFINITION Pestivirus type 1, complete genome.

ACCESSION AF268278

VERSION AF268278.1 GI:9049956

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 12734)

AUTHORS Lai, V.C., Zhong, W., Skelton, A., Ingravallo, P., Vassilev, V.,

Donis, R.O., Hong, Z. and Lau, J.Y.

TITLE Generation and characterization of a hepatitis C virus NS3

JOURNAL protease-dependent bovine viral diarrhoea virus

MEDLINE J. Virol. 74 (14), 6339-6347 (2000)

REFERENCE 2 (bases 1 to 12734)

AUTHORS Lai, V.C.H. and Hong, Z.

TITLE Direct Submission

JOURNAL Submitted (16-MAY-2000) Antiviral Therapy, Schering-Plough Research

INSTITUTE, 2015 Galloping Hill Road, Kenilworth, NJ 07033-0539, USA

FEATURES Location/Qualifiers

source 1..12734

/organism="Pestivirus type 1"

/mol\_type="genomic RNA"

/db\_xref="taxon:11099"

5'UTR 1..385

CDS 386..12508

/codon\_start=1

/product="polyprotein"

/protein\_id="AAF82566.1"

/db\_xref="GI:9049957"

translation="MELNTNEGSSVIVGRIVLSSGSSITACAAQOTRGLGCKITSLD  
TGRKNOVEGEVQIVSTATQTFLATCINGVCTVYHGAGTRTITASPGPVIOMVNDV  
QDLGWPAPOGSRSLTPTCTGSSDLYLTVRHANVIVPVRRRGDSRGSLLSPRLSYLKG  
SSGGLLCPAGHAGVLPRAAVCTRGVAKAVDFIPVENLETTRSGSGAOTEDVYCCSM  
SYSDTKEGATKKTATQPKDRLEGRKMKIVPKESEKSKTPPDATIVVEGKTVQVRKK  
GKTKSKNTQDLGHNNKPKQESRKKLEKALLAWAIIAIVLFQVTMGENTITQWNLQNG  
TEGIRAMFORGVNRLSHGIPKEICTGVFSLHATDIETLTHGMDASAKTNTYTCR  
LQREHMKHGMWNIENPVLNNTQANLTQDPPRECAVTCRYDRASDLNAVTTQA  
RDSPTPLTGCKGKNFSAFAGILMRGPCNFAIASDVLFEKHERISMFQDTTLYLVGL  
TNSLEGAROGTKLTWLGKOLGILGKLENKSKTWEGAVASPYCDVDRKIGYIAYT  
XNCTPACPLRNTKIYVPGKFDNADCKILHEWGSHLSEVLLSLVLSVDFAPETASV  
MYLLHFSIPSHVDVMDCKTQNLNLTADVTPGSVNNLGNKWCIRPNMWPJET  
TVWLAFEVSQVXVLRALRDLTRIWAATTTAFILCLVKIYRGQMVOQILMLLIT  
GVQGLDCKPEFSAIAKDERIGQGAELTTTKEYSPGKLEDVTMIAMCEDGKLM  
YLQCTRETRYLAIALTRALPTSVFKKDFGRKQEDVEMDNFEGFLCPDCAKPIV  
RGKENTLLNGPAFQWCPICGTGTSCTSFNMDTLATTVRTYRBSKPPHQQGCIIT  
QKNLGEDLHNCILGGNMTCPGQOLLYKGSIESCKWCGQFRESGLPHYPGKGL  
ENETGYRLVDSSTCNREGVAIVPQGLTKRIGKIGTIVQIYQVIAWTKLPGMPCRPETIIS  
LVVVVALIGGRVVLMLVLYWVLSQKALGIQYGSVYVANGNLLTHNNLEVYTFLL  
LYLLREESVKKVLLYHILVWHPISKVITVILMIGDVVKADSGOGEYLGKIDLCFT  
TVLVILGLIARBDPTIVPLVTIMARVTELTQPCQVDIAVAVMTIILNYSYVD  
YFRYKWLQCLISLVSGVFLIRSLIYLGRIEMPEVTIPNWRPLTLLLYLISITVTR  
WKVDVAGLLQCVPIILLVTLTWMADFTLLIPLTVELVKLYLTKVTRTIDERSWLG  
IDTVRDSIYDVDSGEGVIFPSPKQAQGNFSILLPLIKATLISCVSSKWLIIYMSY  
LTLDFMYHMRKVTEESGGTNIISRLVAALIELNMSSEESKGLKFFLLSRLRN  
LIINKRVNETVASHYGEVYGNPKIWTIIKASTLSKRHCIICTVCGEKRWKGTCTC





```
FEATURES      Location/Qualifiers
source        1..78
BASE COUNT    18 a /organism="unknown"
ORIGIN        11 c 18 g 31 t

Alignment Scores:
Pred. No.:    0.0426      Length:    78
Score:        48.00      Matches:    10
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 90.91%      Mismatches: 0
Query Match:  97.96%      Indels:    0
DB:           6          Gaps:      0

US-09-965-594-26 (1-11) x AR145197 (1-78)
Oy      1 GlySerValValIleValGlyArgIleValLeu 11
Db      13 GGTTCGTGTTATTGTTGTAAGATTATTTTA 45

Search completed: August 31, 2003, 00:46:50
Job time : 151.976 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:13:57 : Search time 10.2149 Seconds  
(without alignments)  
2906.924 Million cell updates/sec

Title: US-09-965-594-26  
Perfect score: 49  
Sequence: 1 GSVVIVGRIVL 11

Scoring table: BLOSUM62  
Ygapop 10.0 , Xgapopt 0.5  
Ygapop 10.0 , Ygapopt 0.5  
Fgapop 6.0 , Fgapopt 7.0  
Delop 6.0 , Delopt 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+\_p2n\_model -DEV=xlp  
-O=/cgn2.1/USPTO\_Spool03 -OMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-DB=N\_Geneseq\_19Jun03 -OMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -LOCAL=200 -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594 -CGCN\_1\_1\_1412 -runat\_29082003\_151918\_28302 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03:.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	49	100.0	68	21	AAA73336	HCV NS4A-NS3 fusio
2	49	100.0	71	21	AAA73337	HCV NS4A-NS3 fusio
3	49	100.0	588	21	AAA73328	Hepatitis C virus
4	49	100.0	588	21	AAA73329	Hepatitis C virus
5	49	100.0	618	13	AAQ32477	HCV NS2-NS4 clone
6	49	100.0	12734	24	ABA95615	Chimeric BVDV/HCV
7	48	98.0	75	20	AAJ80310	HCV NS4A-NS3 compl
8	48	98.0	75	20	AAJ80314	HCV NS4A-NS3 compl
9	48	98.0	78	20	AAJ80306	HCV NS4A-NS3 compl
10	48	98.0	78	20	AAJ80290	HCV NS4A-NS3 compl
11	48	98.0	96	20	AAJ80338	HCV NS4A-NS3 compl
12	48	98.0	96	20	AAJ80340	HCV NS4A-NS3 compl
13	48	98.0	161	24	AB235821	Hepatitis C virus
14	48	98.0	161	24	ABX10064	HCV NS4A DNA fragm
15	48	98.0	161	24	ABV78245	Hepatitis C virus
16	48	98.0	161	24	ABL91786	HCV polynucleotide
17	48	98.0	189	15	AAQ58472	HCV peptide C14-1
18	48	98.0	267	15	AAQ58473	HCV peptide C14-1
19	48	98.0	279	17	AAT26969	HCV II chimeric ep
20	48	98.0	283	18	AAT49363	Hepatitis C virus
21	48	98.0	321	17	AAT26968	HCV I chimeric epi
22	48	98.0	342	13	AAQ23456	DNA encoding non-A
23	48	98.0	372	13	AAQ24561	NANBH peptide B
24	48	98.0	403	13	AAQ25743	Non-A, Non-B Hepat
25	48	98.0	403	13	AAQ25753	Non-A, Non-B Hepat
26	48	98.0	582	15	AAQ62690	HCV antigen. Synt
27	48	98.0	585	21	AA250045	DNA encoding Hepat
28	48	98.0	586	13	AAQ26990	HCV gene 10. Hepa
29	48	98.0	588	14	AAV05564	DNA associated wit
30	48	98.0	612	25	ABX15706	Anti-viral synthe
31	48	98.0	648	20	AAJ80362	HCV NS4A-NS3 compl
32	48	98.0	648	20	AAJ80363	HCV NS4A-NS3 compl
33	48	98.0	648	20	AAJ80365	HCV NS4A-NS3 compl
34	48	98.0	650	20	AAJ80346	HCV NS4A-NS3 compl
35	48	98.0	650	20	AAJ80347	HCV NS4A-NS3 compl
36	48	98.0	651	20	AAJ80342	HCV NS4A-NS3 compl
37	48	98.0	651	20	AAJ80343	HCV NS4A-NS3 compl
38	48	98.0	651	20	AAJ80344	HCV NS4A-NS3 compl
39	48	98.0	651	20	AAJ80345	HCV NS4A-NS3 compl
40	48	98.0	651	20	AAJ80348	HCV NS4A-NS3 compl
41	48	98.0	651	20	AAJ80349	HCV NS4A-NS3 compl
42	48	98.0	651	20	AAJ80350	HCV NS4A-NS3 compl
43	48	98.0	651	20	AAJ80351	HCV NS4A-NS3 compl
44	48	98.0	654	21	AA250043	DNA encoding hepat
45	48	98.0	669	13	AAQ27012	HK10. Hepatitis C

ALIGNMENTS

RESULT 1  
AAA73336  
ID AAA73336 standard; DNA; 68 BP.  
XX  
AC AAA73336;  
XX  
DT 19-DEC-2000 (first entry)  
XX  
DE HCV NS4A-NS3 fusion protease oligonucleotide #1.  
XX  
KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; primer; ss.  
XX  
OS Synthetic.  
XX  
PN WO200040707-A1.  
XX

PD 13-JUL-2000.  
 XX  
 PF 06-JAN-2000; 2000WO-US00345.  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 XX  
 XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Example 2; Page 22; 66pp; English.  
 XX  
 CC The present sequence is one of two oligonucleotides coding for NS4A and  
 CC linker segments which were used in the creation of a fusion molecule of  
 CC the coding sequences for Hepatitis C virus (HCV) NS3 and NS4A protease  
 CC enzymes. These proteins are both essential for the replication of the  
 CC virus, acting to cleave its replicative proteins from the polyprotein  
 CC produced from the HCV genome. Inhibitors of the two proteins should be  
 CC effective as antiviral treatments of HCV infection. This is useful as HCV  
 CC can lead to chronic liver disease such as cirrhosis, liver failure and  
 CC liver cancer. The present invention concerns a number of NS3 mutants and  
 CC NS3-NS4A fusion proteins which can be used to identify inhibitors of this  
 CC type, as well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.  
 XX  
 SQ Sequence 68 BP; 20 A; 14 C; 17 G; 17 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.106 Length: 68  
 Score: 49.00 Matches: 11  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73336 (1-68)

QY 1 GlySerValIleValGlyArgIleValIleu 11  
 DB 14 GGATCCGTTGTTATCGTCGCCGTATAGTACTG 46

RESULT 2  
 AAA73337/C  
 ID AAA73337 standard; DNA; 71 BP.  
 XX  
 AC AAA73337;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE HCV NS4A-NS3 fusion protease oligonucleotide #2.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; primer; ss.  
 OS  
 OS Synthetic.  
 XX  
 XX WO200040707-A1.  
 PN  
 XX  
 PD 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 PF  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX  
 DR WPI; 2000-465976/40.  
 XX  
 PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
 PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
 PT amino acid, useful for screening inhibitors that may treat hepatitis C  
 PT  
 XX  
 PS Example 2; Page 23; 66pp; English.  
 XX  
 CC The present sequence is one of two oligonucleotides coding for NS4A and  
 CC linker segments which were used in the creation of a fusion molecule of  
 CC the coding sequences for Hepatitis C virus (HCV) NS3 and NS4A protease  
 CC enzymes. These proteins are both essential for the replication of the  
 CC virus, acting to cleave its replicative proteins from the polyprotein  
 CC produced from the HCV genome. Inhibitors of the two proteins should be  
 CC effective as antiviral treatments of HCV infection. This is useful as HCV  
 CC can lead to chronic liver disease such as cirrhosis, liver failure and  
 CC liver cancer. The present invention concerns a number of NS3 mutants and  
 CC NS3-NS4A fusion proteins which can be used to identify inhibitors of this  
 CC type, as well as enabling structural studies of the protease and  
 CC protease:inhibitor complexes.  
 XX  
 SQ Sequence 71 BP; 17 A; 18 C; 15 G; 21 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.111 Length: 71  
 Score: 49.00 Matches: 11  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73337 (1-71)

QY 1 GlySerValIleValGlyArgIleValIleu 11  
 DB 60 GGATCCGTTGTTATCGTCGCCGTATAGTACTG 28

RESULT 3  
 AAA73328  
 ID AAA73328 standard; DNA; 588 BP.  
 XX  
 AC AAA73328;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Hepatitis C virus NS4A-NS3 fusion protease coding sequence #1.  
 XX  
 KW Hepatitis; NS3 protease; viral replication; chronic liver disease;  
 KW liver failure; liver cancer; ds.  
 XX  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 1..588  
 FT /\*tag- a  
 FT /product- "NS3-NS4A fusion protein"  
 XX  
 XX WO200040707-A1.  
 PN  
 XX  
 PD 13-JUL-2000.  
 XX  
 XX 06-JAN-2000; 2000WO-US00345.  
 PF  
 XX  
 PR 08-JAN-1999; 99US-0115271.  
 XX  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Wittekind M, Weinheimer S, Zhang Y, Goldfarb V;  
 XX

DR WPI: 2000-465976/40.  
DR P-PSDB; AAB15212.

XX Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT -

XX Claim 26; Fig 10; 66pp; English.

PS Disclosure; Fig 10; 66pp; English.

XX The present sequence is the coding sequence for a fusion protein created  
CC using the Hepatitis C virus (HCV) NS3 and NS4A protease enzymes. These  
CC proteins are both essential for the replication of the virus, acting to  
CC cleave its replicative proteins from the polyprotein produced from the  
CC HCV genome. Inhibitors of the two proteins should be effective as  
CC antiviral treatments of HCV infection. This is useful as HCV can lead to  
CC chronic liver disease such as cirrhosis, liver failure and liver cancer.  
CC The present invention concerns a number of NS3 mutants and NS3-NS4A  
CC fusion proteins which can be used to identify inhibitors of this type, as  
CC well as enabling structural studies of the protease and  
CC protease-inhibitor complexes.

XX Sequence 588 BP; 97 A; 183 C; 153 G; 155 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 1.21 Length: 588  
Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73328 (1-588)

Qy 1 GlySerValIleValGlyArgIleValLeu 11  
Db 13 GGTCGGTGTATCGTCGCCGTATAGTACTG 45

RESULT 4

AAA73329  
ID AAA73329 standard; DNA; 588 BP.

XX AAA73329;

XX 19-DEC-2000 (first entry)

XX Hepatitis C virus NS4A-NS3 fusion protease coding sequence #2.

XX Hepatitis; NS3 protease; viral replication; chronic liver disease;  
KW liver failure; liver cancer; mutant; mutain; ds.

XX Hepatitis C virus.

OS Synthetic.

PH Key Location/Qualifiers

FT CDS 1..588

FT /\*Lag= a

FT /product= "NS4A-NS3 fusion protein #2"

XX WO200040707-A1.

XX 13-JUL-2000.

XX 06-JAN-2000; 2000WO-US00345.

XX 08-JAN-1999; 99US-0115271.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Wittekand M, Weinheimer S, Zhang Y, Goldfarb V;

XX WPI: 2000-465976/40.

XX P-PSDB; AAB15220.

XX

PT Modified hepatitis C virus (HCV) NS3 protease comprising at least 1  
PT substitution of a hydrophobic alpha-helix 0 amino acid to a hydrophilic  
PT amino acid, useful for screening inhibitors that may treat hepatitis C  
PT -

XX Claim 26; Fig 12; 66pp; English.

XX The present sequence is the coding sequence for a mutated version of a  
CC fusion protein created using the Hepatitis C virus (HCV) NS3 and NS4A  
CC protease enzymes. These proteins are both essential for the replication  
CC of the virus, acting to cleave its replicative proteins from the  
CC polyprotein produced from the HCV genome. Inhibitors of the two proteins  
CC should be effective as antiviral treatments of HCV infection. This is  
CC useful as HCV can lead to chronic liver disease such as cirrhosis, liver  
CC failure and liver cancer. The present invention concerns a number of NS3  
CC mutants and NS3-NS4A fusion proteins which can be used to identify  
CC inhibitors of this type, as well as enabling structural studies of the  
CC protease and protease-inhibitor complexes. The protein produced from this  
CC sequence contains the alpha-helix0-1 variant.

XX Sequence 588 BP; 103 A; 180 C; 156 G; 149 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 1.21 Length: 588  
Score: 49.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-965-594-26 (1-11) x AAA73329 (1-588)

Qy 1 GlySerValIleValGlyArgIleValLeu 11  
Db 13 GGATCGGTGTATCGTCGCCGTATAGTACTG 45

RESULT 5

AAA32477

ID AAA32477 standard; DNA; 618 BP.

XX AAA32477;

XX 25-MAR-2003 (updated)

XX 26-APR-1993 (first entry)

XX HCV NS2-NS4 clone N13-1.

XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;  
KW transcriptase; cDNA; primer; allele; ss.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT CDS 9..608

FT /\*tag= a

XX EP518313-A2.

XX 16-DEC-1992.

XX 11-JUN-1992; 92EP-0109812.

XX 11-JUN-1991; 91JP-0139268.

XX 12-JUL-1991; 91JP-0172794.

XX 07-OCT-1991; 91JP-0287008.

XX 16-DEC-1991; 91JP-0332329.

XX 20-APR-1992; 92JP-0099957.

XX (MITU ) MITSUBISHI KASEI CORP.

XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

XX Teranishi Y;

XX WPI; 1992-417213/51.  
 DR P-FSDB; AAR29846.  
 XX  
 PT New hepatitis C virus gene and its encoded protein - used for  
 PT diagnosing and vaccinating against hepatitis C virus infections  
 XX  
 PS Disclosure; Page 125-26; 305pp; English.  
 XX  
 CC The sequences given in AAQ32472-82 and AAQ32442 are various clones which  
 CC encode the NS2-NS4 regions of the hepatitis C virus (HCV) gene of  
 CC the invention. These sequences were isolated from the serum of a  
 CC patient suffering from hepatitis C (HC). The isolated RNA sequences  
 CC were converted into cDNA using transcriptase in the presence of one  
 CC of the primer sequences given in AAQ32553-64. The sequences were  
 CC then amplified using primer pairs. The cDNA sequences isolated  
 CC represent different alleles of the same region of the HCV gene.  
 CC Sequence comparisons of these clones showed that it is possible for a  
 CC patient to carry more than one HCV strain at one time. See also  
 CC AAQ32436.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 618 BP; 135 A; 184 C; 176 G; 123 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1.28 Length: 618  
 Score: 49.00 Matches: 11  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 13 Gaps: 0  
 US-09-965-594-26 (1-11) x AAQ32477 (1-618)  
 QY 1 GlySerValIleValGlyArgIleValIleu 11  
 Db 351 GGCACGCTGGTCATTTGGGGCAGGATCGTCTTG 383  
 RESULT 6  
 AAQ95615  
 ID ABA95615 standard; DNA; 12734 BP.  
 AC ABA95615;  
 XX  
 DT 21-MAR-2002 (first entry)  
 DE Chimeric BYDV/HCV NS3-wt sequence.  
 XX  
 DE Pestivirus; Npro; protease; NS3; screening; ds.  
 KW  
 XX Chimeric - Bovine viral diarrhea virus.  
 OS Chimeric - Hepatitis C virus.  
 XX  
 PN US6326137-B1.  
 XX  
 PD 04-DEC-2001.  
 XX  
 PF 25-JUN-1999; 99US-0344456.  
 XX  
 PR 25-JUN-1999; 99US-0344456.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Hong Z, Lai VCH, Lau JYN;  
 XX  
 DR WPI; 2002-121103/16.  
 XX  
 CC Nucleic acid construct encoding chimeric Hepatitis C Virus (HCV)  
 PT pestivirus genome where the Npro protease gene is replaced with NS3  
 PT protease gene, useful for in vivo screening of compounds which inhibit  
 PT HCV infection -  
 XX  
 PS Example 2; Columns 17-28; 20pp; English.

XX The present invention relates to a nucleic acid construct encoding a  
 CC chimeric Hepatitis C virus (HCV)-pestivirus genome. The construct  
 CC comprises a pestivirus genome where a Npro pestivirus protease gene is  
 CC replaced with a gene encoding a functional HCV NS3 protease. Furthermore,  
 CC each junction site recognised by the Npro protease is replaced with a  
 CC junction site recognised by the HCV NS3 protease. The construct is useful  
 CC for screening compounds that inhibit HCV in vivo by inhibiting HCV  
 CC protease, where screening may be in cell culture or in an animal model.  
 CC The present sequence is a chimeric clone of BYDV (bovine viral diarrhea  
 CC virus)/HCV NS3-wt, which was used to illustrate the present invention.  
 XX  
 SQ Sequence 12734 BP; 4032 A; 2604 C; 3295 G; 2803 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 39.4 Length: 12734  
 Score: 49.00 Matches: 11  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-965-594-26 (1-11) x ABA95615 (1-12734)  
 QY 1 GlySerValIleValGlyArgIleValIleu 11  
 Db 413 GGTAGTGTGTATTGTGTAGAAATGTTTAA 445  
 RESULT 7  
 AAQ80310  
 ID AAX80310 standard; DNA; 75 BP.  
 XX  
 AC AAX80310;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 DE HCV NS4A-NS3 complex construction primer SEQ ID NO:48.  
 XX  
 KW HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.  
 XX  
 OS Synthetic.  
 OS Hepatitis C virus.  
 XX  
 PN W0928482-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 24-NOV-1998; 98WO-US24528.  
 XX  
 PR 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX  
 PA (SCHE ) SCHERING CORP.  
 XX  
 PI Malcolm BA, Taremi SS, Weber PC, Yao N;  
 XX  
 DR WPI; 1999-385385/32.  
 XX  
 PT New hepatitis C virus covalent complexes  
 XX  
 PS Example 1; Page 32; 21pp; English.  
 XX  
 CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent

CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.

SQ Sequence 75 BP; 16 A; 14 C; 15 G; 30 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.186 Length: 75  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80310 (1-75)

OY 1 GlySerValIleValGlyArgIleValIleu 11  
 DB 13 GGTCTGTGTATTGTTGGTAGAATTATTTTA 45

#### RESULT 8

AAX80314  
 ID AAX80314 standard; DNA; 75 BP.

XX AAX80314;  
 XX 07-SEP-1999 (first entry)  
 XX HCV NS4A-NS3 complex construction primer SEQ ID NO:55.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.  
 OS Hepatitis C virus.  
 PN WO9928482-A2.  
 XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.  
 XX 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 1; Page 33; 21pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.

SQ Sequence 75 BP; 16 A; 12 C; 16 G; 31 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.186 Length: 75  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80314 (1-75)

OY 1 GlySerValIleValGlyArgIleValIleu 11  
 DB 13 GGTCTGTGTATTGTTGGTAGAATTATTTTA 45

#### RESULT 9

AAX80306  
 ID AAX80306 standard; DNA; 78 BP.

XX AAX80306;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:42.

DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 KW NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.  
 OS Hepatitis C virus.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

PR 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX Malcolm BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 1; Page 31; 21pp; English.

CC The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.

SQ Sequence 78 BP; 16 A; 14 C; 17 G; 31 T; 0 other;

#### Alignment Scores:

Pred. No.: 0.195 Length: 78  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80306 (1-78)

QY 1 GlySerValIleValGlyArgIleValLeu 11  
 DB 13 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 45

RESULT 10

AAX80290  
 ID AAX80290 standard; DNA; 78 BP.

XX AAX80290;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:26.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.

XX Hepatitis C virus.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX MalcolM BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 1; Page 26; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity.  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.

XX Sequence 78 BP; 18 A; 11 C; 18 G; 31 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.195 Length: 78  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80290 (1-78)

QY 1 GlySerValIleValGlyArgIleValLeu 11  
 DB 13 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 45

RESULT 11

AAX80338  
 ID AAX80338 standard; DNA; 96 BP.

XX AAX80338;

XX 07-SEP-1999 (first entry)

XX HCV NS4A-NS3 complex construction primer SEQ ID NO:87.

XX HCV; hepatitis C virus; single chain recombinant complex; linker;

XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;

XX hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.

XX Synthetic.

XX Hepatitis C virus.

XX WO9928482-A2.

XX 10-JUN-1999.

XX 24-NOV-1998; 98WO-US24528.

XX 28-JUL-1998; 98US-0094331.

XX 28-NOV-1997; 97US-0067315.

XX (SCHE ) SCHERING CORP.

XX MalcolM BA, Taremi SS, Weber PC, Yao N;

XX WPI; 1999-385385/32.

XX New hepatitis C virus covalent complexes

XX Example 2; Page 43; 21lpp; English.

XX The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity.  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.

XX Sequence 96 BP; 21 A; 17 C; 21 G; 37 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.246 Length: 96  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80338 (1-96)

QY 1 GlySerValIleValGlyArgIleValLeu 11  
 DB 31 GGTCTGTTGTTATTGTTGTTAGAAATTATTTTA 63

RESULT 12

AAX80340  
 ID AAX80340 standard; DNA; 96 BP.

XX AAX80340;

XX 07-SEP-1999 (first entry)



XX HCV NS4A-NS3 complex construction primer SEQ ID NO:89.  
 DE HCV; hepatitis C virus; single chain recombinant complex; linker;  
 XX NS3 protease; NS4A cofactor; NS4A-NS3 complex; serine protease;  
 KW hydrophobic domain; covalent complex; detection; inhibitor; primer; ss.  
 KW Synthetic.  
 OS Hepatitis C virus.  
 OS WO9928482-A2.  
 PN 10-JUN-1999.  
 XX 24-NOV-1998; 98WO-US24528.  
 PF 28-JUL-1998; 98US-0094331.  
 PR 28-NOV-1997; 97US-0067315.  
 XX (SCHE ) SCHERING CORP.  
 PA Malcolm BA, Taremi SS, Weber PC, Yao N;  
 PI WPI: 1999-385385/32.  
 XX New hepatitis C virus covalent complexes  
 PT Example 2; Page 44; 21pp; English.  
 PS The present invention describes a covalent hepatitis C virus (HCV)  
 CC NS4A-NS3 complex comprising a central hydrophobic domain of native HCV  
 CC NS4A peptide, a linker, and an HCV NS3 serine protease domain, where the  
 CC hydrophobic domain of native HCV NS4A peptide is tethered by the linker  
 CC to the amino terminus of the HCV NS3 protease domain. The covalent  
 CC NS4A-NS3 complexes are useful for structural determination and  
 CC determination of mode of binding of HCV inhibitors by NMR spectroscopy.  
 CC They can also be used for detecting inhibitors of the protease activity,  
 CC the helicase activity and the ATPase activity of NS3. The covalent  
 CC NS4A-NS3 complexes are more soluble, stable and active than the  
 CC non-covalent protease-peptide complexes previously available. The  
 CC present sequence represents a primer used in the construction of the  
 CC HCV NS4A-NS3 complexes.  
 XX SQ Sequence 96 BP; 21 A; 17 C; 21 G; 37 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.246 Length: 96  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 20 Gaps: 0

US-09-965-594-26 (1-11) x AAX80340 (1-96)  
 QY 1 GlySerValIleValGlyArgIleValLeu 11  
 DB 31 GGTTCTGTTGTTATGCTGCTAGAAATATTTTA 63

RESULT 13  
 AB235821  
 ID AB235821 standard; DNA; 161 BP.  
 XX AC AB235821;  
 XX 07-FEB-2003 (first entry)  
 XX Hepatitis C virus NS4A polynucleotide SEQ ID NO 129.  
 KW Double stranded RNA; dsRNA; RNAi; RNA inhibition; cytostatic; virucide;  
 KW protozoicide; gene expression; antisense; tumour; infection; Plasmodium;  
 KW virus; viroid; anti-GFP; human; HTV; human immunodeficiency virus;  
 KW Hepatitis C virus; human papilloma virus; gene; ds.

XX Hepatitis C Virus.  
 OS DE10100588-A1.  
 PN 18-JUL-2002.  
 XX 09-JAN-2001; 2001DE-1000588.  
 XX 09-JAN-2001; 2001DE-1000588.  
 PR (RIBO-) RIBOPHARMA AG.  
 PA Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 PI WPI: 2002-683450/74.  
 XX Inhibiting expression of target genes, useful e.g. for treating tumors,  
 PT by introducing into cells two double-stranded RNAs that are  
 PT complementary to the target  
 XX Claim 13; Page 87; 100pp; German.  
 XX The invention relates to inhibiting expression of a target gene in a cell  
 CC by introducing at least two oligoribonucleotides (dsRNAi and II), both  
 CC with a double-stranded (ds) structure of at most 49 sequential nucleotide  
 CC pairs. At least part of one strand (S1, S2) of the ds structures in each  
 CC of dsRNAi and II are complementary to regions in the target gene. The  
 CC method uses antisense inhibition of gene expression using double stranded  
 CC RNA inhibition (RNAi). The method is particularly used to treat tumours  
 CC or infections, especially by Plasmodium or viruses/viroids (pathogenic on  
 CC humans, animals or plants). The method provides more effective inhibition  
 CC of expression than known methods using a single dsRNA, even at very low  
 CC concentrations. When dsRNA has at least one unpaired nucleotide at the  
 CC end, stability (and thus effective concentration in the cell) is  
 CC improved and efficiency can be increased further by pretreating the cells  
 CC with interferon. The present sequence is that of a target DNA of the  
 CC invention.  
 XX SQ Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.442 Length: 161  
 Score: 48.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 90.91% Mismatches: 0  
 Query Match: 97.96% Indels: 0  
 DB: 24 Gaps: 0

US-09-965-594-26 (1-11) x AB235821 (1-161)  
 QY 1 GlySerValIleValGlyArgIleValLeu 11  
 DB 60 GGCAGCGTGTGTCATTGTGGCAGGATCATCTTG 92

RESULT 14  
 ABX10064  
 ID ABX10064 standard; DNA; 161 BP.  
 XX AC ABX10064;  
 XX 23-JAN-2003 (first entry)  
 XX HCV NS4A DNA fragment SEQ ID 129.  
 DE Oligoribonucleotide; interferon; oncogene; cytokine; Id; developmental;  
 KW prion; inhibition; ds.  
 XX Hepatitis C virus.  
 OS DE10100587-C1.  
 PN 21-NOV-2002.

XX 09-JAN-2001; 2001DE-1000587.  
 PF 09-JAN-2001; 2001DE-1000587.  
 XX 09-JAN-2001; 2001DE-1000587.  
 XX (RIBO-) RIBOPHARMA AG.  
 XX Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 XX WPI; 2002-742209/81.  
 XX Inhibiting expression of target genes, e.g. oncogenes, in cells, by  
 PT introduction of complementary double-stranded oligoribonucleotide,  
 PT after treating the cell with interferon  
 XX Disclosure; Page 92; 98pp; German.  
 XX This invention describes a novel method for inhibiting expression of a  
 CC target gene by introducing into the cell that contains the target gene  
 CC at least one oligoribonucleotide (dsRNAi) that has a double-stranded  
 CC (ds) structure of not more than 49 consecutive nucleotides (nt), where  
 CC at least a segment of one strand of the ds structure is complementary  
 CC with the target gene and the cells are treated with interferon before  
 CC introduction of dsRNAi. The method is used to inhibit expression of  
 CC target genes, particularly oncogenes, cytokine genes, Id (not defined)  
 CC protein genes, developmental or prion genes, or genes expressed in  
 CC pathogenic organisms (particularly plasmidia) or in viruses or viroids  
 CC (pathogenic in humans, animals or plants). Treating the cells with  
 CC interferon greatly increases the extent to which dsRNA can inhibit  
 CC expression of the target genes, and the effect is even greater when dsRNA  
 CC are modified to increase their stability. ABX09936-ABX10075 represent  
 CC gene fragments used to illustrate the method of the invention.  
 XX SQ Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other:  
 Alignment Scores: Length: 161  
 Pred. No.: 0.442 Matches: 10  
 Score: 48.00 Conservative: 1  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 90.91% Indels: 0  
 Query Match: 97.96% Gaps: 0  
 DB: 24  
 US-09-965-594-26 (1-11) x ABX10064 (1-161)  
 Qy 1 GlySerValValIleValGlyArgIleValIleu 11  
 |||||  
 Db 60 GGCACGGTGGTCATTGTGGCAGCATCTTIG 92  
 RESULT 15  
 ABV78245  
 ID ABV78245 standard; DNA; 161 BP.  
 XX AC ABV78245;  
 XX 15-NOV-2002 (first entry)  
 XX Hepatitis C virus NS4A DNA SEQ ID NO 129.  
 DE RNA inhibition; dsRNAi; gene expression inhibitor; oncogene; cytostatic;  
 KW virucide; protozoacide; gene; ds.  
 XX Hepatitis C virus.  
 OS WO200255693-A2.  
 PN 18-JUL-2002.  
 PD 09-JAN-2002; 2002WO-EP00152.  
 XX 09-JAN-2001; 2001DE-1000586.  
 PR 26-OCT-2001; 2001DE-1055280.  
 PR 29-NOV-2001; 2001DE-1056411.

PR 07-DEC-2001; 2001DE-1060151.  
 XX (RIBO-) RIBOPHARMA AG.  
 XX Kreutzer R, Limmer S, Rost S, Hadwiger P;  
 XX WPI; 2002-590671/63.  
 XX Inhibiting expression of target gene, useful e.g. for inhibiting  
 PT oncogenes, by administering double-stranded RNA complementary to the  
 PT target and having an overhang  
 XX Claim 10; Page 190; 203pp; German.  
 XX The invention relates to inhibiting expression of a target gene (I) in a  
 CC cell by introducing an inhibitory RNA (dsRNAi) having a double-stranded  
 CC structure of at most 49 consecutive bases. At least part of one strand  
 CC (asI) of dsRNAi is complementary to (I) and at least one end of dsRNAi  
 CC has an overhang of 1-4 nucleotides. The method is used to inhibit the  
 CC expression of a wide range of genes, e.g. oncogenes, cytokine genes etc.  
 CC in humans, also genes in plasmidium or in viruses or viroids that are  
 CC pathogenic for humans, animals or plants. Introducing an overhang into  
 CC dsRNA greatly increases effectiveness for inhibiting gene expression,  
 CC both in vivo and in vitro and also increases stability and thus the  
 CC effective concentration inside the cell. The present sequence is that of  
 CC a gene related to the invention.  
 XX SQ Sequence 161 BP; 32 A; 40 C; 55 G; 34 T; 0 other:  
 Alignment Scores: Length: 161  
 Pred. No.: 0.442 Matches: 10  
 Score: 48.00 Conservative: 1  
 Percent Similarity: 100.00% Mismatches: 0  
 Best Local Similarity: 90.91% Indels: 0  
 Query Match: 97.96% Gaps: 0  
 DB: 24  
 US-09-965-594-26 (1-11) x ABV78245 (1-161)  
 Qy 1 GlySerValValIleValGlyArgIleValIleu 11  
 |||||  
 Db 60 GGCACGGTGGTCATTGTGGCAGCATCTTIG 92

Search completed: August 30, 2003, 19:48:23  
 Job time : 12.2149 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 30, 2003, 19:20:43 ; Search time 106.667 Seconds  
(without alignments)  
2506.388 Million cell updates/sec

Title: US-09-965-594-26  
Perfect score: 49  
Sequence: 1 GSVVIVGRIVL 11

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09965594/runat\_29082003\_151919\_28322/app.query.fasta\_1.2872  
-DB=EST -OPT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=pto -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09965594.ecgn\_1\_1\_12630\_@runat\_29082003\_151919\_28322 -NCPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	91.8	234	9	AW580055
2	43	87.8	618	14	CA728563
3	43	87.8	664	13	BM300987
4	43	87.8	709	13	BM306038
5	43	87.8	725	13	BM176544
6	42	85.7	150	10	BE069681
7	42	85.7	646	10	BE069684
8	42	85.7	658	13	BQ121075
9	42	85.7	777	10	BG000305
10	42	85.7	1035	13	BUT18639
11	41	83.7	222	9	AU183349
12	41	83.7	487	12	B1882963
13	41	83.7	502	10	BE556909
14	41	83.7	564	14	CB417286
15	41	83.7	584	12	B1981053
16	41	83.7	597	12	B1840872
17	41	83.7	603	12	BM185756
18	41	83.7	607	13	BQ093551
19	41	83.7	620	12	B1982002
20	41	83.7	629	10	BE016243
21	41	83.7	641	10	BE016238
22	41	83.7	641	10	BE201405
23	41	83.7	660	14	CD392599
24	41	83.7	676	9	AW019436
25	41	83.7	796	9	AF122168
26	41	83.7	882	13	BQ734390
27	41	83.7	886	12	B1759235
28	41	83.7	1569	29	CC250893
29	40	81.6	315	9	AV211292
30	40	81.6	398	29	CNS03X0L
31	40	81.6	540	10	BG403954
32	40	81.6	667	14	CB030392
33	40	81.6	693	28	BZ004158
34	40	81.6	771	13	BQ785063
35	40	81.6	791	29	AG096061
36	40	81.6	1280	29	AG092509
37	40	81.6	1657	10	BG708760
38	40	81.6	1976	11	BC035952
39	39	79.6	116	10	BE069567
40	39	79.6	135	10	BE069567
41	39	79.6	170	10	BE071133
42	39	79.6	178	10	BE069730
43	39	79.6	195	10	BE071108
44	39	79.6	234	9	AV012254
45	39	79.6	251	10	BB441428

ALIGNMENTS

RESULT 1  
AW580055  
LOCUS RC1-HT0375-130100-011-905 HT0375 Homo sapiens cDNA, mRNA linear EST 16-MAR-2000  
DEFINITION AW580055  
ACCESSION AW580055  
VERSION AW580055.1 GI:7255104  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 234)  
AUTHORS HGCP <http://www.ludwig.org.br/ORESTES>.

TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-RC1&l2-RC1-HT0375-130100-011-g05&t3=2000-01-13&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 28  
High quality sequence stop: 216.

## FEATURES

source  
1..234  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0375"

/note="Organ: head/neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 52 a 35 c 76 g 71 t

## ORIGIN

Alignment Scores: 26.4 Length: 234  
Pred. No.: 45.00 Matches: 8  
Score: 100.00% Conservative: 3  
Percent Similarity: 72.73% Mismatches: 0  
Best Local Similarity: 91.84% Indels: 0  
Query Match: 9 Gaps: 0  
DB:

US-09-965-594-26 (1-11) x AW580055 (1-234)

QY 1 GlySerValIleValGlyArgIleValLeu 11  
|||||:|||||:|||||:|||||:|||||

Db 75 GGTAGTATTATATTCTGGGACGATTAGTTT 107

## RESULT 2

CA728563/c

LOCUS wdlc.pk004.f22 wdlc Triticum aestivum cDNA clone wdlc.pk004.f22  
DEFINITION 5' end, mRNA sequence.

ACCESSION CA728563.1 GI:25450552

VERSION EST.

KEYWORDS Triticum aestivum (bread wheat)

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; 1 Triticeae; Triticum.

REFERENCE 1 (bases 1 to 618)

AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,

Miao,G., Caraher,N. and Hanafey,M.K.

TITLE Dupont Wheat cDNA Sequence

JOURNAL Unpublished

COMMENT Contact: Scott V. Tingey

Crop Genetics

E. I. Dupont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

## FEATURES

## source

Seq primer: M13.  
Location/Qualifiers  
1..618  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_wdlc.pk004.f22"  
/tissue\_type="inflorescence"  
/lab\_host="DH10B"  
/clone\_lib="wdlc"

/note="Vector: pBluescript SK+; Site\_1: EcoRI; Site\_2: XhoI; Wheat (Triticum aestivum, Hi Line) developing inflorescence +/- 4 cm"

BASE COUNT 140 a 194 c 168 g 110 t 6 others

## ORIGIN

Alignment Scores: 235 Length: 618  
Pred. No.: 43.00 Matches: 8  
Score: 90.91% Conservative: 2  
Percent Similarity: 72.73% Mismatches: 1  
Best Local Similarity: 87.76% Indels: 0  
Query Match: 14 Gaps: 0  
DB:

US-09-965-594-26 (1-11) x CA728563 (1-618)

QY 1 GlySerValIleValGlyArgIleValLeu 11

|||||:|||||:|||||:|||||:|||||

Db 113 GGAGGGGTGCTGCTGGCAGGCTGTCCTT 81

## RESULT 3

BW300987

## LOCUS

DEFINITION BW300987 Nori Satoh unpublished cDNA library, linear EST 11-NOV-2002  
Intestinalis cDNA clone cinc020g06 5', mRNA sequence.

ACCESSION BW300987

VERSION BW300987.1 GI:24881598

KEYWORDS EST.

SOURCE Clona Intestinalis

ORGANISM Clona Intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Clona.

REFERENCE 1 (bases 1 to 664)

AUTHORS Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.

TITLE Expressed genes in Clona Intestinalis (2002c)

JOURNAL Unpublished

COMMENT Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satohe@scidian.zool.kyoto-u.ac.jp.

## FEATURES

## source

Location/Qualifiers  
1..664  
/organism="Clona Intestinalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:7719"  
/clone\_wdlc.cinc020g06"  
/tissue\_type="neural complex"  
/clone\_lib="Nori Satoh unpublished cDNA library, neural complex"

BASE COUNT 201 a 145 c 135 g 183 t

## ORIGIN

Alignment Scores: 259 Length: 664  
Pred. No.: 43.00 Matches: 9  
Score: 100.00% Conservative: 1  
Percent Similarity: 90.00% Mismatches: 0  
Best Local Similarity: 87.76% Indels: 0  
Query Match: 13 Gaps: 0  
DB:

US-09-965-594-26 (1-11) x BW300987 (1-664)

Qy 1 GlySerValValIleValGlyArgIleVal 10  
 |||||  
 Db 418 GGAAGTGTGTGTGTGTGTGAAGATTGTT 447

## RESULT 4

BW306038

LOCUS

DEFINITION BW306038 Nori Satoh unpublished cDNA library, heart EST 11-NOV-2002  
 intestinalis cDNA clone ciht012f15 5', mRNA sequence.

ACCESSION BW306038

VERSION 1 GI:24886649

KEYWORDS EST.

SOURCE

ORGANISM

Ciona intestinalis  
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Clonidae; Ciona.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satohe@acidian.zool.kyoto-u.ac.jp.

FEATURES

source

1..709

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="ciht012f15"

/tissue\_type="heart"

/clone\_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT

ORIGIN

199 a 156 c 155 g 199 t

Alignment Scores:

Pred. No.: 283

Score: 43.00

Percent Similarity: 100.00%

Best Local Similarity: 90.00%

Query Match: 87.76%

DB: 13

Length: 709

Matches: 9

Conservative: 1

Mismatches: 0

Indels: 0

Gaps: 0

US-09-965-594-26 (1-11) x BW306038 (1-709)

Qy 1 GlySerValValIleValGlyArgIleVal 10

|||||

Db 307 GGAAGTGTGTGTGTGTGAAGATTGTT 336

RESULT 5

BW176544/c

LOCUS

DEFINITION BW176544 Nori Satoh unpublished cDNA library, heart EST 04-NOV-2002  
 intestinalis cDNA clone rciht012f15 3', mRNA sequence.

ACCESSION BW176544

VERSION 1 GI:24566468

KEYWORDS EST.

SOURCE

ORGANISM

Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Clonidae; Ciona.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satohe@acidian.zool.kyoto-u.ac.jp.

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-4081  
 Fax: 81-75-705-1113  
 Email: satohe@acidian.zool.kyoto-u.ac.jp.

## FEATURES

source

1..725

/organism="Ciona intestinalis"

/mol\_type="mRNA"

/db\_xref="taxon:7719"

/clone="rciht012f15"

/tissue\_type="heart"

/clone\_lib="Nori Satoh unpublished cDNA library, heart"

BASE COUNT

ORIGIN

195 a 168 c 155 g 207 t

Alignment Scores:

Pred. No.: 292

Score: 43.00

Percent Similarity: 100.00%

Best Local Similarity: 90.00%

Query Match: 87.76%

DB: 13

Length: 725

Matches: 9

Conservative: 1

Mismatches: 0

Indels: 0

Gaps: 0

US-09-965-594-26 (1-11) x BW176544 (1-725)

Qy 1 GlySerValValIleValGlyArgIleVal 10

|||||

Db 546 GGAAGTGTGTGTGTGTGAAGATTGTT 517

RESULT 6

BE069681

LOCUS

DEFINITION RC2-BT0389-120400-014-c02 BT0389 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE069681

VERSION BE069681.1 GI:8414331

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 150)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-RC2-BT0389-120>)

400-014-c02&t3=2000-04-12&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 3

High quality sequence stop: 150.

Location/Qualifiers

1..150

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

FEATURES

source

```

/clone.lib="BT0389"
/Note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      38 a      18 c      45 g      49 t
ORIGIN
Alignment Scores:
Pred. No.:      55.3      Length:      150
Score:          42.00     Matches:      8
Percent Similarity: 90.91% Conservative: 2
Best Local Similarity: 72.73% Mismatches: 1
Query Match:     85.71% Indels:      0
DB:             10      Gaps:      0

US-09-965-594-26 (1-11) x BE069681 (1-150)
QY      1 GlySerValValIleValGlyArgIleValLeu 11
Db      12 GGTAGTACTAATAATTGTGGGAGATTAGTTT 44

RESULT 7
LOCUS      BG590884      646 bp      mRNA      linear      EST 07-MAR-2003
DEFINITION EST498726 P. infestans-challenged leaf Solanum tuberosum cDNA clone
BPL16H5 5' sequence, mRNA sequence.
ACCESSION  BG590884
VERSION     BG590884.1 GI:13609024
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 646)
Zhang,P., Hernandez,M., Tornqvist,C.-E., Wirtz,U., Loukianov,A.,
Rangel,P., Haberland,G.T., Cho,J., Chiemiango,A., Bougri,O., Buell
,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
Generation of ESTs from Potato Leaves Challenged with Phytophthora
infestans, Incompatible Reaction
Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1..646
                     .organism="Solanum tuberosum"
                     .mol_type="mRNA"
                     .cultivar="Kennebec"
                     .db_xref="taxon:4113"
                     .clone="BPL16H5"
                     .tissue_type="leaf"
                     .dev_stage="6 week old"
                     .lab_host="SOLR"
                     .clone_lib="P. infestans-challenged leaf"
                     .note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Whole plants were challenged with 450,000
Sporangia/ml P. infestans US-1(US 940501) in Biotron
(Madison, Wisconsin). Leaf tissue was collected at 1, 2,
5, 12, and 24 hours post-challenge and frozen in liquid
nitrogen immediately upon removal. Kennebec plants showed
no signs of HR. Katahdin plants (susceptible to P.
infestans US-1) were used as controls and showed
infection. NOTE: We cannot exclude the possibility that

```

```

this sequence is actually derived from Phytophthora rather
than potato."
BASE COUNT      159 a      110 c      163 g      214 t
ORIGIN
Alignment Scores:
Pred. No.:      390      Length:      646
Score:          42.00     Matches:      9
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 81.82% Mismatches: 1
Query Match:     85.71% Indels:      0
DB:             10      Gaps:      0

US-09-965-594-26 (1-11) x BG590884 (1-646)
QY      1 GlySerValValIleValGlyArgIleValLeu 11
Db      117 GGTCTGGTAGTATAGTGGGAGAAATCATAC 149

RESULT 8
LOCUS      BQ121075      658 bp      mRNA      linear      EST 07-MAR-2003
DEFINITION EST606651 mixed potato tissues Solanum tuberosum cDNA clone STMEV29
5' end, mRNA sequence.
ACCESSION  BQ121075
VERSION     BQ121075.2 GI:21920306
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 658)
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karanycheva,S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1..658
                     .organism="Solanum tuberosum"
                     .mol_type="mRNA"
                     .cultivar="Kennebec or Binjte"
                     .db_xref="taxon:4113"
                     .clone="STMEV29"
                     .tissue_type="mixed tissues"
                     .lab_host="SOLR"
                     .clone_lib="Mixed potato tissues"
                     .note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT      159 a      97 c      182 g      220 t
ORIGIN
Alignment Scores:
Pred. No.:      399      Length:      658
Score:          42.00     Matches:      9
Percent Similarity: 90.91% Conservative: 1
Best Local Similarity: 81.82% Mismatches: 1
Query Match:     85.71% Indels:      0
DB:             13      Gaps:      0

```



## FEATURES

## source

## Location/Qualifiers

1. .222  
 /organism="Cyprinus carpio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7962"  
 /clone="H7"  
 /tissue\_type="head kidney"  
 /clone\_lib="Cyprinus carpio head kidney stimulated by  
 lipo-polysaccharide and concanavalin-A"  
 /note="common name: common carp ; stimulated by  
 lipo-polysaccharide and concanavalin-A"

## BASE COUNT

73 a 34 c 45 g 70 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 146 Length: 222  
 Score: 41.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 4  
 Best Local Similarity: 63.64% Mismatches: 0  
 Query Match: 83.67% Indels: 0  
 DB: 9 Gaps: 0

US-09-965-594-26 (1-11) x AU183349 (1-222)

QY 1 GlySerValValIleValGlyArgIleValLeu 11

Db 42 GGGCAGTGGTTCATCGGAGAGAGTTCTACTG 74

## RESULT 12

## BI882963/C

## LOCUS

BI882963 f01e01.x1 zebrafish Research Genetics C32 fin Danio rerio cDNA  
 clone IMAGE:4469065 3', mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

BI882963 487 bp mRNA linear EST 16-SEP-2002  
 f01e01.x1 zebrafish Research Genetics C32 fin Danio rerio cDNA  
 clone IMAGE:4469065 3', mRNA sequence.

QY 1 GlySerValValIleValGlyArgIleValLeu 11

Db 42 GGGCAGTGGTTCATCGGAGAGAGTTCTACTG 74

RESULT 12

BI882963/C

LOCUS

BI882963

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

## FEATURES

## source

## Location/Qualifiers

1. .487  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:4469065"  
 /tissue\_type="Fin"  
 /lab\_host="GeneHogs (HS996, a phage-resistant isolate of  
 DH10B)"  
 /clone\_lib="zebrafish Research Genetics C32 fin"

/note="vector: pT7T3D-Pac with a modified polylinker;  
 Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was prepared  
 from zebrafish(C32) fin, and was then primed with a Not I  
 - oligo(dT) primer. Double-stranded cDNA was ligated to  
 Eco RI adaptors (Pharmacia), digested with Not I and  
 cloned into the Not I and Eco RI sites of the modified  
 pT7T3 vector. Library is non-normalized. Library was  
 constructed by Ning Wu. NOTE: This clone is available  
 royalty-free through LLNL; contact the IMAGE Consortium  
 (info.llnl.gov) for further information"

BASE COUNT 144 a 115 c 91 g 137 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 417 Length: 487  
 Score: 41.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 4  
 Best Local Similarity: 63.64% Mismatches: 0  
 Query Match: 83.67% Indels: 0  
 DB: 12 Gaps: 0

US-09-965-594-26 (1-11) x BI882963 (1-487)

QY 1 GlySerValValIleValGlyArgIleValLeu 11

Db 165 GGGCAGTGGTTCATCGGAGAGAGTTCTACTG 133

## RESULT 13

## BE556909

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1. .502

## /organism="Danio rerio"

## /mol\_type="mRNA"

## /db\_xref="taxon:7955"

## /tissue\_type="Fin"

/lab\_host="GeneHogs (HS996, a phage-resistant isolate of  
 DH10B)"

## /clone\_lib="zebrafish Research Genetics C32 fin"

/note="vector: pT7T3D-Pac with a modified polylinker;  
 Site 1: EcoRI; Site 2: NotI; 1st strand cDNA was prepared  
 from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I

## from zebrafish(C32) fin, and was then primed with a Not I



- oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is non-normalized. Library was constructed by Ning Wu. NOTE: This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info.llnl.gov) for further information."

BASE COUNT 136 a 100 c 124 g 142 t

ORIGIN

Alignment Scores:

Pred. No.: 434 Length: 502  
Score: 41.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 63.64% Mismatches: 0  
Query Match: 83.67% Indels: 0  
DB: 10 Gaps: 0

US-09-965-594-26 (1-11) x BE556909 (1-502)

QY 1 GlySerValIleValGlyArgIleValLeu 11

Db 355 GGGCAGTGGTTCATGGGAAGATTGACTG 387  
||||:|||||:||||:||||:|||||

RESULT 14

LOCUS

DEFINITION CB417286 564 bp mRNA linear EST 27-MAR-2003  
STR00763 gastrula stage cDNA library Danio rerio cDNA clone CB380  
5' similar to CCAAT/enhancer binding protein beta, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Thisse B  
Institut de Genetique et de Biologie Molculaire et Cellulaire  
CNRS, INSERM, ULP  
1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex  
, France  
Tel: 33 3 88 65 33 60  
Fax: 33 3 88 65 32 01  
Email: thisse@ipmc.u-strasbg.fr  
EST from a cDNA of a gene whose expression is spatially restricted during embryogenesis. We have established its expression pattern on zebrafish embryos from the gastrula stage to 2 days of development. The corresponding data are available on the zebrafish community database at <http://zfinfo.org/cDNA> library preparation: B. Riggleman. DNA Sequencing by: IGBMC sequencing facility. Clone distribution: zebrafish international resource center at the University of Oregon (Institute of Neuroscience, 1254 University of Oregon, Eugene, OR 97403-1254)  
Seq primer: T3 ATTAACCTCACTAAAGGA.

FEATURES  
source  
1..564  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="CB380"  
/dev\_stage="gastrula stage embryos"  
/note="Vector: Lambda Zap; Site\_1: EcoRI; Site\_2: XhoI; Oligo dT cDNA library constructed from RNA pooled from gastrula stage zebrafish embryos"

BASE COUNT 148 a 133 c 155 g 128 t

ORIGIN

Alignment Scores:

Pred. No.: 507 Length: 564  
Score: 41.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 63.64% Mismatches: 0  
Query Match: 83.67% Indels: 0  
DB: 14 Gaps: 0

US-09-965-594-26 (1-11) x CB417286 (1-564)

QY 1 GlySerValIleValGlyArgIleValLeu 11

Db 507 GGGCAGTGGTTCATGGGAAGATTGACTG 539  
||||:|||||:||||:||||:|||||

RESULT 15

LOCUS

DEFINITION BI981053 584 bp mRNA linear EST 26-JUL-2002  
fu38H04.x1 zebrafish adult brain Danio rerio cDNA clone  
IMAGE:5332062 3' similar to TR:097894 097894 CCAAT/ENHANCER BINDING  
PROTEIN BETA ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@wustl.wustl.edu  
CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: [www.resgen.com](http://www.resgen.com)) (email contact: info@resgen.com) and  
Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))  
Seq primer: -40UP  
High quality sequence stop: 422.

FEATURES  
source  
1..584  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:5332062"  
/sex="mixed male and female"  
/tissue\_type="brain"  
/dev\_stage="adult"  
/lab\_host="E. Coli DH10B"  
/clone\_lib="zebrafish adult brain"  
/note="Vector: pZiPlox; Site\_1: NotI; Site\_2: SalI;  
Original library was constructed in lambdaZiPlox. Mass  
excision of the cDNA library was performed to yield  
pZiPlox plasmids. Insert check was done in original  
library."

BASE COUNT 159 a 151 c 118 g 156 t

ORIGIN

Alignment Scores:  
Pred. No.: 531 Length: 584  
Score: 41.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 4  
Best Local Similarity: 63.64% Mismatches: 0  
Query Match: 83.67% Indels: 0  
DB: 12 Gaps: 0

US-09-965-594-26 (1-11) x 81981053 (1-584)

QY 1 GlySerValValIleValGlyArgIleValLeu 11  
DB 171 GGGCGAGTGGTGTGTCATGGGAAGAGTTGTACTG 139

Search completed: August 31, 2003, 04:27:56  
Job time : 112.667 secs